

Washington University School of Medicine

Digital Commons@Becker

---

Open Access Publications

---

2007

## Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways

Francesca Capotosti  
*University of Lausanne*

James J.-D. Hsieh  
*Washington University School of Medicine in St. Louis*

Winship Herr  
*University of Lausanne*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Please let us know how this document benefits you.**

---

### Recommended Citation

Capotosti, Francesca; Hsieh, James J.-D.; and Herr, Winship, "Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways." *Molecular and Cellular Biology*. 27, 20. 7063-7072. (2007).

[https://digitalcommons.wustl.edu/open\\_access\\_pubs/2269](https://digitalcommons.wustl.edu/open_access_pubs/2269)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [vanam@wustl.edu](mailto:vanam@wustl.edu).

## Species Selectivity of Mixed-Lineage Leukemia/Trithorax and HCF Proteolytic Maturation Pathways

Francesca Capotosti, James J.-D. Hsieh and Winship Herr  
*Mol. Cell. Biol.* 2007, 27(20):7063. DOI:  
10.1128/MCB.00769-07.  
Published Ahead of Print 13 August 2007.

---

Updated information and services can be found at:  
<http://mcb.asm.org/content/27/20/7063>

---

*These include:*

### REFERENCES

This article cites 38 articles, 18 of which can be accessed free at: <http://mcb.asm.org/content/27/20/7063#ref-list-1>

### CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), [more»](#)

---

---

Information about commercial reprint orders: <http://journals.asm.org/site/misc/reprints.xhtml>  
To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

---

## Species Selectivity of Mixed-Lineage Leukemia/Trithorax and HCF Proteolytic Maturation Pathways<sup>∇</sup>

Francesca Capotosti,<sup>1</sup> James J.-D. Hsieh,<sup>2</sup> and Winship Herr<sup>1\*</sup>

Center for Integrative Genomics, University of Lausanne, G  nopode Building, 1015 Lausanne, Switzerland,<sup>1</sup> and Molecular Oncology, Department of Medicine, Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri 63110<sup>2</sup>

Received 2 May 2007/Returned for modification 15 June 2007/Accepted 27 July 2007

Site-specific proteolytic processing plays important roles in the regulation of cellular activities. The histone modification activity of the human trithorax group mixed-lineage leukemia (MLL) protein and the cell cycle regulatory activity of the cell proliferation factor herpes simplex virus host cell factor 1 (HCF-1) are stimulated by cleavage of precursors that generates stable heterodimeric complexes. MLL is processed by a protease called *taspase 1*, whereas the precise mechanisms of HCF-1 maturation are unclear, although they are known to depend on a series of sequence repeats called HCF-1<sub>PRO</sub> repeats. We demonstrate here that the *Drosophila* homologs of MLL and HCF-1, called Trithorax and dHCF, are both cleaved by *Drosophila* *taspase 1*. Although highly related, the human and *Drosophila* *taspase 1* proteins display cognate species specificity. Thus, human *taspase 1* preferentially cleaves MLL and *Drosophila* *taspase 1* preferentially cleaves Trithorax, consistent with coevolution of *taspase 1* and MLL/Trithorax proteins. HCF proteins display even greater species-specific divergence in processing: whereas dHCF is cleaved by the *Drosophila* *taspase 1*, human and mouse HCF-1 maturation is *taspase 1* independent. Instead, human and *Xenopus* HCF-1<sub>PRO</sub> repeats are cleaved in vitro by a human proteolytic activity with novel properties. Thus, from insects to humans, HCF proteins have conserved proteolytic maturation but evolved different mechanisms.

Site-specific proteolysis has emerged as an important mechanism contributing to the regulation of basic cellular processes such as development (24), metabolism (3), cell cycle progression (7), and apoptosis (22). Site-specific proteolysis regulates both physiological and disease processes, often by activating latent functions.

In human cells, a number of nuclear proteins, such as the trithorax group mixed-lineage leukemia (MLL) protein (11, 21, 35) and herpes simplex virus (HSV) host cell factor 1 (HCF-1) (15, 31, 32), are synthesized as precursors that undergo proteolytic maturation to generate stable, noncovalently associated heterodimeric complexes. The importance of these proteolytic maturation processes is underscored by the finding that the *Drosophila* homologs of these proteins, Trithorax (Trx) and dHCF, also undergo proteolytic maturation (16, 18). Nevertheless, the cellular mechanism of proteolytic maturation is completely known only for human MLL.

MLL is cleaved by a novel endopeptidase called *taspase 1* that utilizes an N-terminal threonine generated by autoproteolysis as the nucleophile for polypeptide cleavage (10). For HCF-1, autocatalytic processing of bacterially synthesized HCF-1 has been observed (28), but the mechanisms of HCF-1 maturation in human cells remain to be clarified. Curiously, in addition to sharing proteolytic maturation processes, MLL and HCF-1 bind each other (36), and both play important roles in the regulation of the cell division cycle (6, 23, 26, 30). These relationships encouraged us to compare their proteolytic maturation pathways along with those of the *Drosophila* Trx and dHCF homologs.

Of these four proteins, MLL is the largest, consisting of 3,969 amino acids. It was originally discovered because the gene encoding MLL is the site of chromosomal translocations in human childhood leukemias (1, 5, 8). MLL possesses histone H3 lysine 4 (H3K4) methyltransferase activity (20, 21) and is required for maintaining proper *Hox* and *Cyclin* gene expression (26, 37). MLL is cleaved by *taspase 1* (10) at two sites to generate the associated amino-terminal (MLL<sub>N</sub>) and carboxy-terminal (MLL<sub>C</sub>) subunits (11, 21, 35). This proteolysis enhances the H3K4 methyltransferase activity of the MLL<sub>C</sub> subunit, which promotes *Cyclin* gene expression and cell cycle progression (26).

The enzyme responsible for Trx cleavage is not known, but it has been hypothesized that it is a homolog of *taspase 1* because the region of cleavage (16) contains a putative *taspase 1* recognition site (10, 35). The possible importance of Trx cleavage for its biological function has been indicated by the activity of a mutant Trx protein, called Trx<sup>E3</sup>, which contains a 271-amino-acid deletion (19) that spans the predicted processing site and abrogates Trx protein maturation (16). Trx is required to maintain proper expression of *antennapedia* and *bithorax* complex genes during fly development (2). Interestingly, *trx*<sup>E3</sup> mutants display defective *antennapedia* but not *bithorax* complex gene expression (25), suggesting a selective role of Trx cleavage in its function.

HCF-1 is a 2,035-amino-acid chromatin-associated protein that was first discovered as a transcriptional coactivator for HSV immediate-early gene transcription, where it stimulates formation of the viral VP16-induced transcriptional regulatory complex (see reference 33 for a review). Proteolytic maturation of HCF-1 involves multiple cleavages at any one of six

\* Corresponding author. Mailing address: Center for Integrative Genomics, University of Lausanne, G  nopode Building, 1015 Lausanne, Switzerland. Phone: 41-21-692-39-22. Fax: 41-21-692-39-25. E-mail: winship.herr@unil.ch.

<sup>∇</sup> Published ahead of print on 13 August 2007.

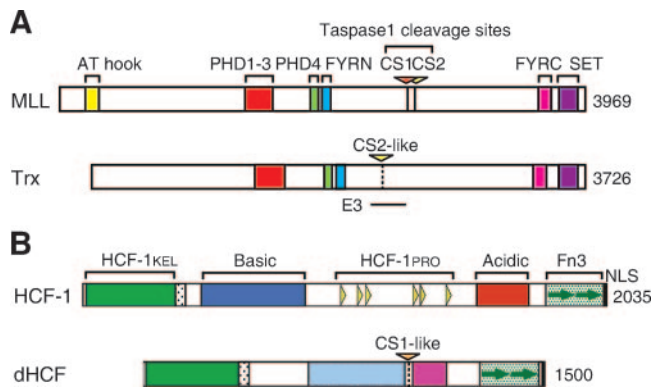


FIG. 1. Human and *Drosophila* MLL/Trx and HCF proteins. (A) Schematic structures of human MLL and *Drosophila* Trx proteins. Architectural elements are identified above the schematic: PHD, plant homeodomain; FYRN, MLL<sub>N</sub> carboxy-terminal association element; FYRC, MLL<sub>C</sub> amino-terminal association element; SET, histone H3 lysine 4 methyltransferase domain. The positions of the two taspase 1 cleavage sites (CS1 and CS2) are indicated as red and yellow arrowheads, respectively. Conserved regions in Trx are shown as for MLL. The line labeled E3 indicates the region deleted in the mutant Trx<sup>E3</sup> protein. (B) Schematic structures of human HCF-1 and dHCF proteins. Architectural elements are identified above the schematic. HCF-1<sub>KEL</sub>, Kelch repeat domain; Basic and Acidic, regions enriched in basic and acidic residues, respectively; HCF-1<sub>PRO</sub>, HCF-1 proteolytic processing repeats; Fn3, fibronectin type 3 repeats; NLS, nuclear localization signal. Conserved regions in dHCF are shown in the same colors; basic and acidic regions that display similar charge bias but not sequence identity are shown in related colors. The position of the dHCF CS1-like taspase 1 recognition site is indicated by the arrowhead and dashed line.

centrally located 26-amino-acid repeats called HCF-1<sub>PRO</sub> proteolytic (HCF-1<sub>PRO</sub>) repeats, generating the associated amino-terminal (HCF-1<sub>N</sub>) and carboxy-terminal (HCF-1<sub>C</sub>) subunits (15, 31, 32). The HCF-1<sub>N</sub> and HCF-1<sub>C</sub> subunits play separate roles in two key phases of the cell cycle: the HCF-1<sub>N</sub> subunit promotes passage through the G<sub>1</sub> phase, and the HCF-1<sub>C</sub> subunit is necessary for proper mitosis and cytokinesis during the M phase (14). Proteolytic processing is necessary to ensure proper HCF-1 function, as HCF-1<sub>C</sub> subunit functions are inhibited if the HCF-1 precursor cannot be processed (14).

Consistent with the importance of HCF-1 proteolytic maturation, the 1,500-amino-acid dHCF protein, although lacking HCF-1<sub>PRO</sub> repeats, also undergoes proteolytic maturation to generate associated dHCF<sub>N</sub> and dHCF<sub>C</sub> subunits (18). The dHCF<sub>N</sub> and dHCF<sub>C</sub> subunits display considerable structural similarity to the human HCF-1<sub>N</sub> and HCF-1<sub>C</sub> subunits (Fig. 1) as well as functional conservation. Thus, as for the HCF-1<sub>N</sub> subunit, the dHCF<sub>N</sub> subunit can associate with the HSV transactivator VP16 and stabilize the VP16-induced transcriptional regulatory complex (18), as well as associate with the *Drosophila* cell cycle regulators dE2F1 (the homolog of human E2F1) and dE2F2 (the homolog of human E2F4) (27).

Following our interest in HCF protein proteolytic maturation and in the conservation of HCF proteins in animals, we noted that dHCF, albeit not human HCF-1, contains a potential taspase 1 cleavage site within the region predicted by Mahajan et al. (18) for cleavage. This observation has led us to find that the dHCF and Trx proteins are both cleaved by the *Drosophila* taspase 1 homolog, whereas human HCF-1 proteo-

lytic cleavage is independent of taspase 1. Thus, the *Drosophila* and human trithorax group proteins share proteolytic maturation pathways, while the *Drosophila* and human HCF proteins, although conserving proteolysis as a maturation process, mature via different proteolytic mechanisms.

## MATERIALS AND METHODS

**Plasmid constructions and DNA template preparations.** A full-length *Drosophila* taspase 1 cDNA (LD05057) was obtained from the *Drosophila* Genomics Resource Center, Bloomington, IN, and cloned into the pET-28a(+) His-tagged bacterial expression vector (Novagen). The His-tagged full-length human taspase 1 cDNA bacterial expression vector and the vector for in vitro transcription/translation containing wild-type and CS1/2-mutated MLL amino acids 2400 to 2900 have been described previously (10). The full-length dHCF cDNA cloned in a T7 epitope-tagged *Drosophila* expression vector (pACXT-dHCF) was a kind gift of Angus Wilson (New York University). For in vitro transcription/translation of the precursor, a PCR fragment encoding dHCF amino acids 733 to 1212 was amplified from pACXT-dHCF and inserted in the pNCITE vector (30). For the Trx precursor, a PCR fragment encoding Trx amino acids 1973 to 2458 was amplified by reverse transcription-PCR from SL2 cell RNA, verified by sequence analysis, and inserted directly into the pGEM-T vector (Promega). Fragments consisting of HCF-1 amino acids 686 to 1166 derived from either wild-type or mutant pCGNHCF<sub>FL</sub> (31, 32) were PCR amplified using a forward primer including the phage T7 promoter and β-globin translational initiation codon and directly used for in vitro transcription/translation. The *Xenopus* HCF-1 (xHCF-1) precursor, corresponding to amino acids 1358 to 1653, was PCR amplified from pCMV-SPORT6-xHCF-1<sub>1358-1653</sub> (clone ID 7653830; BioCat) as for HCF-1 and directly used for in vitro transcription/translation. In vitro transcription/translation and [<sup>35</sup>S]methionine labeling were performed using the TNT T7 Quick transcription/translation system (Promega) as recommended by the manufacturer. Mutants of the Trx CS2-like and dHCF CS1-like sites were generated by QuickChange site-directed mutagenesis (Stratagene).

**Taspase 1 protein expression.** His-tagged human taspase 1 and *Drosophila* taspase 1 were expressed in *Escherichia coli* BL21(DE5) cells and purified by Ni affinity chromatography (QIAGEN) as recommended by the manufacturer. Proteins eluted from the resin were dialyzed against phosphate-buffered saline, 25% glycerol.

**Cell culture and extracts.** HeLa cells were grown at 37°C in Dulbecco's modified Eagle's medium with 10% fetal bovine serum (FBS). *Drosophila* SL2 cells were grown at 25°C in Schneider's medium (GIBCO) with 10% heat-inactivated FBS. Nuclear and cytosolic extracts from HeLa and SL2 cells were prepared as previously described (4). Wild-type or taspase 1<sup>-/-</sup> mouse embryonic fibroblasts (MEFs) from day 12.5 embryos were grown as described previously (26). Extract was prepared by lysis in radioimmunoprecipitation assay buffer with Complete protease inhibitor cocktail (Roche) on ice for 30 min and clarified by centrifugation at 20,000 × g for 20 min.

**In vitro cleavage assays.** Taspase 1 in vitro cleavage assays were performed as previously described (10). For comparative experiments, recombinant human taspase 1 and *Drosophila* taspase 1 activities were titrated on the respective cognate MLL and Trx substrates and used in the amount required to cleave 50% of the substrate (corresponding to 10 ng of human taspase 1 and 50 ng of *Drosophila* taspase 1). For the xHCF-1 in vitro cleavage assay, 1 μg of human taspase 1 and *Drosophila* taspase 1 were used. Cell extract in vitro cleavage assays were performed using 22 μl of the indicated cell extract in a 30-μl reaction mixture and incubating reaction mixtures at 30°C for 8 h or the indicated time period. HeLa cell extract heat treatment was at 65°C for 20 min, and protease inhibitors were utilized at the final concentration of 0.5 mg/ml Pefabloc (Roche) and 1× Complete protease inhibitor cocktail (Roche). The reaction mixtures were resolved by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the resulting radioactive signals were visualized with a Typhoon Trio+ imaging system and quantified with ImageQuant (Amersham Biosciences).

**RNAi and cell transfection.** A 183-bp double-stranded RNA (dsRNA) (RNAi 1) and a 495-bp dsRNA (RNAi 2) against the *Drosophila* taspase 1 gene sequence as well as a nonspecific firefly luciferase dsRNA were used for RNA interference (RNAi). dsRNAs were synthesized with a MEGAscript T7 kit (Ambion), using as templates PCR products amplified with primer pairs (RNA: 2), CAGTCGT TCTGTGTCGGCTA and CTAGGAGCTGCAGGAAGGTG (RNA: 1) and TGGGTGTCCTCATGGTG and TCCAGCGTACAGATCCTG each containing an additional 5' T7 promoter sequence. For dsRNA treatment, 1 × 10<sup>6</sup> SL2 cells were seeded in six-well plates in 1 ml of serum-free Schneider's medium and



incubated at 25°C for 2 h in the presence of 15 µg of dsRNA before addition of 2 ml of Schneider's medium with 10% heat-inactivated FBS. dsRNA-treated cells were collected after 96 h. For transient transfection,  $2 \times 10^6$  untreated or 48-h-dsRNA-treated SL2 cells were transfected with 5 µg of DNA using Cellfectin (Invitrogen) as described by the manufacturer. Cells were collected at 48 h posttransfection. For collection, SL2 cells were washed with phosphate-buffered saline, lysed in Laemmli SDS sample buffer, and subsequently analyzed by immunoblot analysis.

**Immunoblot analysis.** Protein lysates were resolved by SDS-PAGE, transferred to nitrocellulose membrane, and probed with T7 tag monoclonal antibody (Novagen), polyclonal HCF-1<sub>C</sub> antiserum (H12) (31), or polyclonal dHCF<sub>N</sub> antiserum, followed by appropriate IRDye800-conjugated anti-mouse or IRDye680-conjugated anti-rabbit immunoglobulin G (Rockland). The immunoreactive bands were detected by fluorescence with an Odyssey infrared imager (LI-COR). The polyclonal dHCF<sub>N</sub> antiserum was raised in rabbits against a 15-amino-acid peptide (EGSDFVDPAFSSGER) corresponding to the N terminus of dHCF and affinity purified with the peptide immunogen.

## RESULTS

Figure 1 shows schematic representations of the human and *Drosophila* MLL/Trx and HCF proteins highlighting conserved sequence and structural elements. Figure 1A also shows the positions of the two known taspase 1 cleavage sites in MLL, called CS1 (QVD/GADD) and CS2 (QLD/GVDD) (10, 11, 35). By sequence similarity, Trx contains a CS2-like site (QMD/GVDD) (10, 11, 35) and dHCF contains a CS1-like site (QLD/GAGD), both of which lie in positions analogous to the cleavage sites in the respective human homologs (i.e., MLL CS1 and CS2 sites and HCF-1<sub>PRO</sub> sites). We therefore tested whether the CS2-like Trx and CS1-like dHCF taspase 1 recognition sites are substrates for taspase 1.

**Trx and dHCF are cleaved by *Drosophila* taspase 1.** In initial experiments, we tested the ability of human taspase 1 to cleave the Trx and dHCF proteins and found that the human protease showed a strong preference for the human MLL substrate (see below). Therefore, we tested the ability of the *Drosophila* taspase 1 homolog to cleave Trx and dHCF substrates as shown in Fig. 2. To facilitate this analysis, we generated for each protein shorter precursor cleavage substrates that contain the regions with the putative taspase 1 cleavage sites. Both wild-type and cleavage site mutant substrates were used (Fig. 2A). These precursors were synthesized *in vitro* in the presence of [<sup>35</sup>S]methionine and incubated with recombinant *Drosophila* taspase 1 purified from *E. coli*, and the products were separated by SDS-PAGE. Figure 2B shows that *Drosophila* taspase 1 can cleave the wild-type Trx precursor (lanes 1 and 2) but not the precursor carrying the mutation in the CS2-like site (lanes 3 and 4). These results suggest that, like its human homolog MLL, Trx is cleaved by taspase 1 at the CS2-like cleavage site. Similarly, as shown in Fig. 2C, *Drosophila* taspase 1 can effectively cleave the wild-type dHCF precursor (lanes 1 and 2) but not the CS1-like mutant precursor (lanes 3 and 4). These experiments define the first heterologous protease for HCF protein maturation, the Trx protease *Drosophila* taspase 1.

To characterize this dHCF protease further, we assessed whether dHCF is cleaved at the taspase 1 cleavage site *in vivo*. We transfected *Drosophila* SL2 cells with an epitope-tagged expression vector encoding full-length wild-type and CS1-like mutant recombinant dHCF (rdHCF) and assayed cleavage by immunoblotting with an antitag antibody. As shown in Fig. 2D, synthesis of wild-type rdHCF resulted in an approximately 105-kDa N-terminal fragment (rdHCF<sub>N</sub>) (lane 2) (18),

whereas synthesis of the CS1-like mutant rdHCF led to an approximately 180-kDa product, a size equivalent to that of the full-length protein (rdHCF<sub>FL</sub>) (lane 3) (18). These results indicate that dHCF is cleaved at a single taspase 1 recognition site.

We also analyzed proteolytic maturation of endogenous dHCF by generating an antibody directed to the N terminus of the protein (see Materials and Methods). Immunoblot analysis of the same SL2 extracts shown in Fig. 2D revealed that endogenous dHCF<sub>N</sub> comigrates with the processed rdHCF<sub>N</sub> (compare lanes 1 and 2 in Fig. 2D and E). With this antibody, we could not detect any endogenous full-length dHCF protein, but we could detect the ectopic CS1-like mutant rdHCF<sub>FL</sub> (compare lane 3 with lane 1). The lack of endogenous full-length dHCF indicates, as previously shown using a dHCF<sub>C</sub> antibody (9), that dHCF is efficiently processed in *Drosophila* cells.

To determine whether *Drosophila* taspase 1 is the sole protease responsible for dHCF processing *in vivo*, we depleted *Drosophila* taspase 1 from *Drosophila* cells by RNAi. As shown in Fig. 2F, independent treatment of SL2 cells with two different *taspase 1* dsRNAs (RNAi 1 and RNAi 2) led in each case to the detection of dHCF<sub>FL</sub> (lanes 3 and 4) compared to untreated (lane 1) or mock RNAi-treated (luciferase, lane 2) cells. The levels of dHCF<sub>FL</sub> were limited, however, possibly owing to high levels of stable dHCF<sub>N</sub> remaining from before the RNAi treatment. In contrast, we observed prominent inhibition of rdHCF cleavage when the RNAi-induced *Drosophila* taspase 1 depletion was associated with concomitant synthesis of dHCF by simultaneous transfection of the epitope-tagged dHCF expression vector. Thus, as shown in Fig. 2G, in *taspase 1* RNAi-treated cells, rdHCF processing is significantly reduced (lanes 3 and 4) compared to that in untreated (lane 1) or mock RNAi-treated (lane 2) cells. In summary, our *in vitro* and *in vivo* results indicate that *Drosophila* taspase 1 is responsible for dHCF processing at a single site and indeed may be the sole protease responsible for dHCF maturation. Thus, in flies, where dHCF lacks HCF-1<sub>PRO</sub> repeats, HCF protein maturation is apparently not autocatalytic as observed for an HCF-1<sub>PRO</sub> repeat region of human HCF-1 synthesized in *E. coli* (28).

**dHCF is not an effective substrate for human taspase 1.** We were surprised to find that dHCF is a substrate of *Drosophila* taspase 1, because Izeta et al. (12) have shown that dHCF is not processed in hamster cells and yet such cells would be expected to possess taspase 1. To explore this apparent discrepancy, we compared directly the abilities of human and *Drosophila* cell extracts to cleave the dHCF precursor. Indeed, as the results of Izeta et al. (12) would suggest, the dHCF precursor, albeit cleaved by a *Drosophila* SL2 extract, was not effectively cleaved by a human HeLa cell extract, as shown in Fig. 3A (compare lanes 2 and 3). This dissimilarity is likely the result of intrinsic differences in the human taspase 1 and *Drosophila* taspase 1 enzymes, because the same species-specific activity was also observed with purified recombinant human taspase 1 and *Drosophila* taspase 1 (compare lanes 5 and 6) (see Materials and Methods for enzymatic activity normalization). These results explain why Izeta et al. (12) did not observe dHCF processing by using mammalian cells and emphasize the

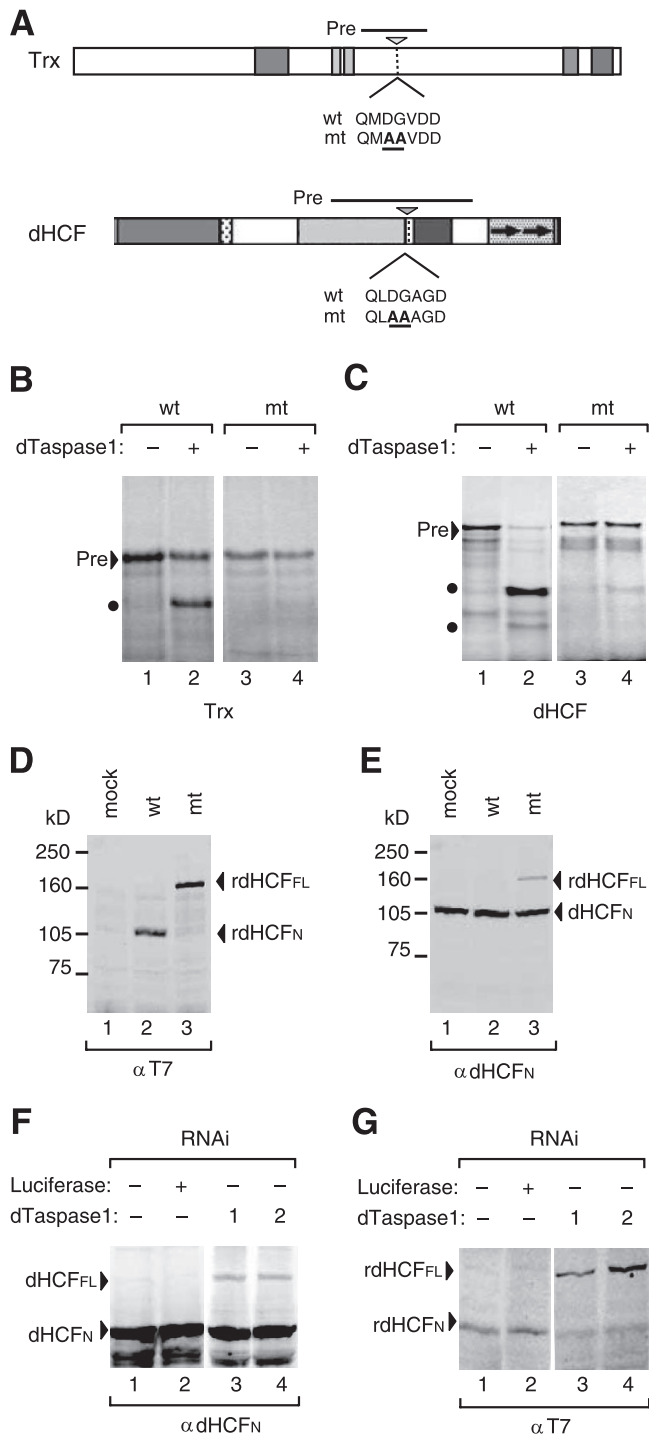


FIG. 2. Trx and dHCF are cleaved by *Drosophila* taspase 1. (A) Schematic of Trx and dHCF cleavage precursors. The lines labeled Pre indicate the region contained within each precursor. Wild-type (wt) and mutant (mt) versions of the putative taspase 1 cleavage sites are shown below. Note that the two diagrams are not drawn to the same scale. (B) *Drosophila* taspase 1 (dTaspase1) proteolytic activity on the Trx precursor. <sup>35</sup>S-labeled wild-type (lanes 1 and 2) or CS2-like mutant (lanes 3 and 4) Trx precursors were incubated for 2 h at 37°C with (lanes 2 and 4) or without (lanes 1 and 3) purified recombinant *Drosophila* taspase 1. Products were resolved by SDS-PAGE and revealed by autoradiography. ●, N-terminal cleavage product. A smaller C-terminal fragment is not visible owing to the reduced specific activity of this product. (C) *Drosophila* taspase 1 proteolytic activity on the

importance of using a homologous system to assay HCF protein processing (18).

**The human and *Drosophila* taspase 1 enzymes have co-evolved with their specific MLL/Trx substrates.** The dHCF *Drosophila* taspase 1 and human taspase 1 cleavage results suggest that taspase 1 activity has changed during evolution. To study this evolution further, we compared the human taspase 1 and *Drosophila* taspase 1 enzymes on homologous substrates: precursors from the human MLL and *Drosophila* Trx proteins (Fig. 3B). As shown in Fig. 3C, a human HeLa cell and a *Drosophila* SL2 cell extract display enhanced activity on the cognate MLL and Trx precursor, respectively (compare lanes 2 and 3 with lanes 5 and 6) (in this assay only MLL-CS2 cleavage is indicated). Furthermore, purified recombinant human taspase 1 and *Drosophila* taspase 1 display the same cognate substrate preference (Fig. 3D, compare lanes 2 and 3 with lanes 5 and 6). These results suggest that the taspase 1 enzymes have coevolved with their substrates MLL and Trx.

Zhou et al. (38) have defined a human taspase 1 recognition heptapeptide sequence. The MLL CS2 and Trx CS2-like sites differ by a single amino acid (from QLDGVDD in MLL CS2 to QMDGVDD in Trx). To examine the determinants responsible for the cognate enzyme substrate preference of *Drosophila* taspase 1 on Trx, we converted the Trx CS2-like taspase 1 site to the MLL CS2 sequence via a single amino acid change (Fig. 3B). As shown in Fig. 3E, the humanized Trx precursor [Trx<sub>(MLLCS2)</sub>] is still a better substrate for *Drosophila* taspase 1 than human taspase 1, indicating that the species-specific taspase 1 selectivity observed here results from sequence differences that lie outside the so-far-identified heptapeptide recognition sequence.

**Human HCF-1 is not a taspase 1 substrate.** HCF-1 is cleaved at any one of the six HCF-1<sub>PRO</sub> repeats (15, 31), and when the region containing all six sites is deleted (14, 32) or the six sites are all individually inactivated by amino acid substitution (29), the protein is no longer cleaved. To study HCF-1

dHCF precursor. <sup>35</sup>S-labeled wild-type (lanes 1 and 2) or CS1-like mutant (lanes 3 and 4) dHCF precursors were incubated with (lanes 2 and 4) or without (lanes 1 and 3) purified recombinant *Drosophila* taspase 1. Products were resolved by SDS-PAGE and revealed by autoradiography. ●, location of larger N-terminal and smaller C-terminal cleavage products. (D) dHCF cleavage at the *Drosophila* taspase 1 cleavage site in vivo. SL2 cells were mock transfected (lane 1) or transfected with wild-type (lane 2) or CS1-like mutant (lane 3) T7-tagged full-length dHCF expression vector. Proteolysis by endogenous protease was assessed by anti-T7 tag (αT7) immunoblotting. rdHCF<sub>FL</sub>, full-length rdHCF; rdHCF<sub>N</sub>, rdHCF N-terminal subunit. Molecular mass markers are listed on the left. (E) Anti-dHCF<sub>N</sub> antibody (αdHCF<sub>N</sub>) reveals processing of endogenous dHCF. Endogenous dHCF proteolysis of the samples shown in panel D was revealed by immunoblot analysis with affinity-purified anti-dHCF<sub>N</sub> antibody. (F) RNAi depletion of *Drosophila* taspase 1 impairs endogenous dHCF processing. SL2 cells were mock treated (lane 1) or treated with luciferase (lane 2) or two independent taspase 1 dsRNAs (RNAi 1 in lane 3 and RNAi 2 in lane 4) for 48 h before cleavage analysis by anti-dHCF<sub>N</sub> immunoblotting. (G) RNAi depletion of *Drosophila* taspase 1 impairs dHCF processing. SL2 cells were treated with the indicated dsRNAs as in panel F for 48 h before transfection of the T7-tagged full-length dHCF expression vector. dHCF cleavage was analyzed 48 h after the transfection by anti-T7 immunoblotting.

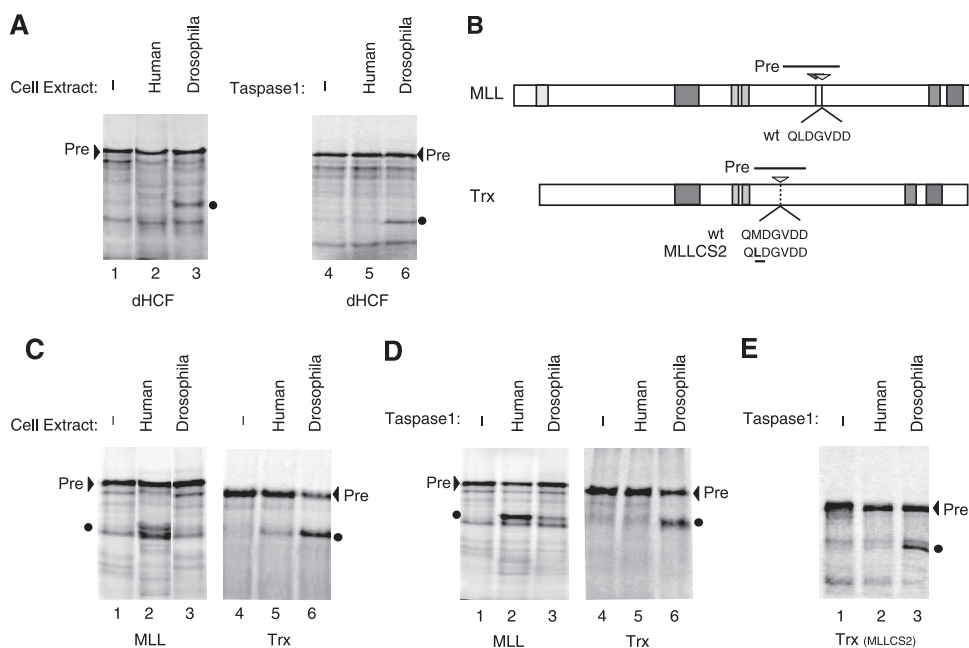


FIG. 3. Coevolution of human and *Drosophila* taspase 1 with their specific substrates. (A) Human taspase 1 versus *Drosophila* taspase 1 proteolytic activity on dHCF cleavage. Wild-type dHCF precursor was incubated without extract (lanes 1 and 4) or with HeLa (lane 2) and SL2 (lane 3) cytosolic extracts or recombinant human taspase 1 (lane 5) and *Drosophila* taspase 1 (lane 6). (B) Schematic of MLL and Trx cleavage precursors. The line labeled Pre indicates the region contained within each precursor. The Trx<sub>(MLL CS2)</sub> precursor contains the Trx cleavage site (QMDGVDD) changed to the MLL CS2 site sequence (QLDGVDD). (C) Human and *Drosophila* cell extract activities on MLL and Trx cleavage. MLL (lanes 1 to 3) and Trx (lanes 4 to 6) precursors were incubated without extract (lanes 1 and 4) or with HeLa (lanes 2 and 5) or SL2 (lanes 3 and 6) cytosolic extracts. (D) Human taspase 1 and *Drosophila* taspase 1 activities on MLL and Trx cleavage. MLL (lanes 1 to 3) and Trx (lanes 4 to 6) precursors were incubated without taspase 1 (lanes 1 and 4) or with recombinant human taspase 1 (lanes 2 and 5) or *Drosophila* taspase 1 (lanes 3 and 6). (E) Human taspase 1 and *Drosophila* taspase 1 activities on humanized Trx<sub>(MLLCS2)</sub> precursor. The Trx<sub>(MLLCS2)</sub> precursor was incubated without taspase 1 (lane 1) or with recombinant human taspase 1 (lane 2) or *Drosophila* taspase 1 (lane 3). For all cleavage products, black dots indicate the N-terminal cleavage products; smaller, lower-specific-activity C-terminal fragments are not visible.

proteolytic cleavage, we first asked whether the HCF-1<sub>PRO</sub> repeat is a taspase 1 substrate. Therefore, we prepared an in vitro HCF-1 protease substrate containing three of the six HCF-1<sub>PRO</sub> repeats (HCF-1<sub>PRO</sub> repeats 1, 2, and 3, called HCF-1rep123) (Fig. 4A) and incubated it with recombinant human taspase 1. Figure 4B shows that human taspase 1 had no effect on this HCF-1 precursor (compare lanes 1 and 2) at a concentration that effectively cleaved the MLL precursor (compare lanes 5 and 6), suggesting that taspase 1 is not an HCF-1<sub>PRO</sub> repeat protease.

To further exclude the requirement of taspase 1 in HCF-1 maturation, we asked whether HCF-1 processing is affected by the absence of taspase 1. Takeda et al. (26) have shown using MEFs that MLL processing is disrupted in *taspase 1*<sup>-/-</sup> knock-out mice. In contrast, as shown in Fig. 4C, HCF-1 is processed normally in both wild-type and *taspase 1*<sup>-/-</sup> MEFs (compare lanes 2 and 3), with the pattern of cleavage products being identical to the one observed in a HeLa cell extract (lane 1). Together, these results indicate that taspase 1 is not the protease responsible for HCF-1<sub>PRO</sub> repeat processing. Indeed, taspase 1 is unlikely to be involved in HCF-1 processing generally, as there is no identifiable taspase 1 cleavage site in the entire HCF-1 amino acid sequence.

**A human cell activity cleaves HCF-1<sub>PRO</sub> repeats.** To identify an activity responsible for HCF-1<sub>PRO</sub> repeat processing, we used a wild-type HCF-1rep123 substrate or substrates with

HCF-1<sub>PRO</sub> repeats 1 and 2 mutated either individually (HCF-1repX23 and HCF-1rep1X3) or together (HCF-1repXX3) as illustrated in Fig. 5A. As shown in Fig. 5B, in the absence of cell extract, none of the precursors was effectively cleaved (lanes 1, 3, 5, and 7). In the presence of HeLa cell extract, however, each precursor was cleaved at the wild-type but not mutated HCF-1<sub>PRO</sub> repeats (compare lanes 2, 4, 6, and 8), displaying a cleavage pattern consistent with HCF-1<sub>PRO</sub> repeat specific cleavage. These results suggest that HeLa cell extracts possess an activity that can specifically cleave an HCF-1<sub>PRO</sub> repeat precursor.

To establish the nature of this HCF-1<sub>PRO</sub> repeat enzymatic activity, we performed a time course experiment using the HCF-1rep123 precursor. As shown in Fig. 5C, under the same assay conditions used for Fig. 5B, the HCF-1<sub>PRO</sub> repeat cleavage product increased with incubation time for approximately 8 h. Accumulation of the product resulting from cleavage at the HCF-1<sub>PRO</sub> repeat 1 was quantified as shown in Fig. 5D. These results suggest that in cell extracts the HCF-1<sub>PRO</sub> repeat proteolytic cleavage is relatively slow but stable for 8 h.

**The HeLa cell HCF-1<sub>PRO</sub> repeat proteolytic activity is heat sensitive but protease inhibitor resistant.** To characterize the HCF-1<sub>PRO</sub> repeat protease further, we assayed the sensitivity of the HeLa cell extract activity to heat and protease inhibitor treatment as shown in Fig. 5E. The HeLa cell activity was sensitive to heat treatment (compare lanes 1 to 3) but resistant



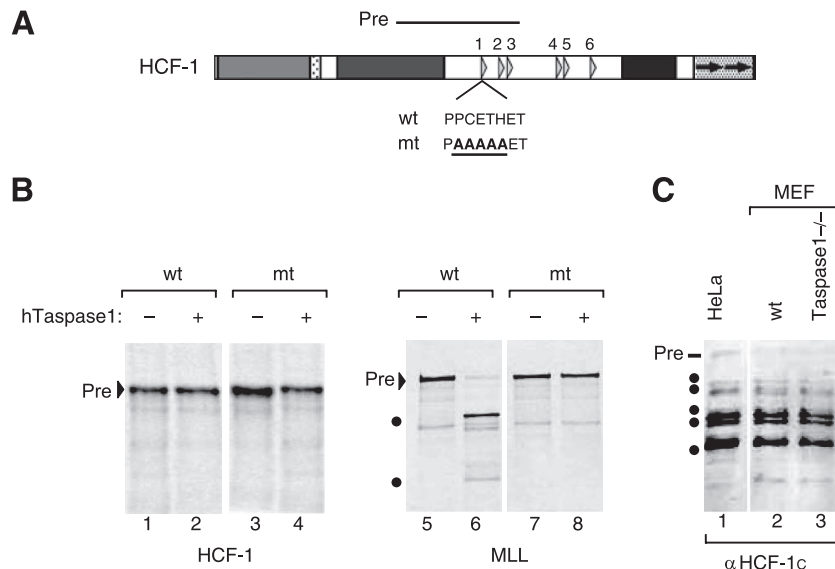


FIG. 4. HCF-1 is not cleaved by taspase 1. (A) Schematic of HCF-1 cleavage precursor. The line labeled Pre indicates the region contained within the precursor. Wild-type (wt) and mutated (mt) versions of the HCF-1<sub>PRO</sub> repeat 1 are shown below. (B) HCF-1<sub>PRO</sub> repeats are not evident substrates for human taspase 1 (hTaspase1) cleavage. <sup>35</sup>S-labeled wild-type (lanes 1, 2, 5, and 6) or mutated (lanes 3, 4, 7, and 8) HCF-1 and MLL precursors were incubated with (lanes 2, 4, 6, and 8) or without (lanes 1, 3, 5, and 7) purified recombinant human taspase 1. ●, larger N-terminal and shorter C-terminal cleavage products. (C) *taspase 1*<sup>-/-</sup> cells contain normally processed HCF-1. Wild-type (lane 2) and *taspase 1*<sup>-/-</sup> (lane 3) MEF lysates were resolved by 7% SDS-PAGE. Cleavage of endogenous HCF-1 was detected by anti-HCF-1<sub>C</sub> (αHCF-1<sub>C</sub>) immunoblot analysis. HeLa nuclear extract (lane 1) was used as an HCF-1 cleavage pattern control. —, HCF-1 precursor; ●, HCF-1<sub>PRO</sub> repeat cleavage products.

to Pefabloc, a serine protease inhibitor (compare lanes 3 and 4) (an assay with thrombin and a substrate with a thrombin cleavage site showed that the serine protease inhibitor was active [data not shown]), and Complete, a serine and cysteine protease inhibitor cocktail (compare lanes 3 and 5). Thus, the HCF-1<sub>PRO</sub> repeat protease displays both sensitivity (heat) and resistance (serine and cysteine protease inhibitors) to various treatments. We note that the resistance of the HCF-1<sub>PRO</sub> repeat protease to a serine protease inhibitor differentiates it from the autocatalytic activity described by Vogel and Kristie, which was serine protease inhibitor sensitive (28).

**HCF protein maturation mechanism has changed during evolution.** The identification of taspase 1 as the protease responsible for dHCF maturation and the observation that human HCF-1 is cleaved by a different activity at the HCF-1<sub>PRO</sub> repeats indicate that, from fly to human, HCF proteins have conserved proteolytic maturation but evolved different mechanisms. These differences between flies and humans appear to be generally specific to insects or vertebrates, as illustrated in Fig. 6. Thus, the *Apis mellifera* (honeybee) HCF protein possesses a consensus taspase 1 cleavage site at the same relative location as in the dHCF protein (Fig. 6A), suggesting that taspase 1 cleavage of HCF proteins may be generally conserved in insects. In parallel, a comparison of human, frog, and fish HCF-1 proteins shows that the position (Fig. 6B) and sequence (Fig. 6C) of the HCF-1<sub>PRO</sub> repeats have been very highly conserved in these three distantly related vertebrate species. Thus, between insects and vertebrates, there appears to have been an evolutionary switch in HCF protein processing mechanism that has been highly conserved within each group.

Interestingly, in some fish (e.g., *F. rubripes*) and frog (e.g., *X. tropicalis*) species, in addition to the perfectly conserved HCF-

1<sub>PRO</sub> repeats there is a partially conserved taspase 1 cleavage site (six-of-seven match to a heptapeptide consensus taspase 1 cleavage site) (38) located just C terminal of the last HCF-1<sub>PRO</sub> repeat (Fig. 6B). This partially conserved taspase 1 cleavage site could represent an active site, indicating that in some species HCF-1 could be processed by both taspase 1 and an HCF-1<sub>PRO</sub> repeat protease. Alternatively, it could be inactive and instead represent the vestige of an active taspase 1 cleavage site in an ancestral HCF-1 protein. To distinguish between these two possibilities, we asked whether an xHCF-1 cleavage precursor containing the partially conserved taspase 1 cleavage site and the adjacent two HCF-1<sub>PRO</sub> repeats (xHCF-1rep89 [Fig. 6D]) could be a substrate for either taspase 1 and/or the HCF-1<sub>PRO</sub> repeat protease. As shown in Fig. 6D, the xHCF-1rep89 substrate is not cleaved by either *Drosophila* taspase 1 or human taspase 1 at concentrations that can effectively cleave the dHCF precursor (compare lanes 1 to 3 and 4 to 6). Instead, as for human HCF-1 (lanes 8 and 9), the xHCF-1rep89 precursor is cleaved by the HeLa cell extract at positions corresponding to the HCF-1<sub>PRO</sub> repeats (compare lanes 4 and 7). These data suggest that, like mammalian HCF-1 proteins, xHCF-1 is not cleaved by taspase 1 but is instead processed at the HCF-1<sub>PRO</sub> repeats by the same activity responsible for human HCF-1 maturation. The ability of the human cell extract to cleave the *Xenopus* HCF-1<sub>PRO</sub> repeats provides functional evidence of the high degree of conservation of the HCF-1<sub>PRO</sub> repeat proteolytic process in vertebrates.

## DISCUSSION

In this study, we have compared the proteolytic maturation processes of two pairs of evolutionarily conserved proteins: the



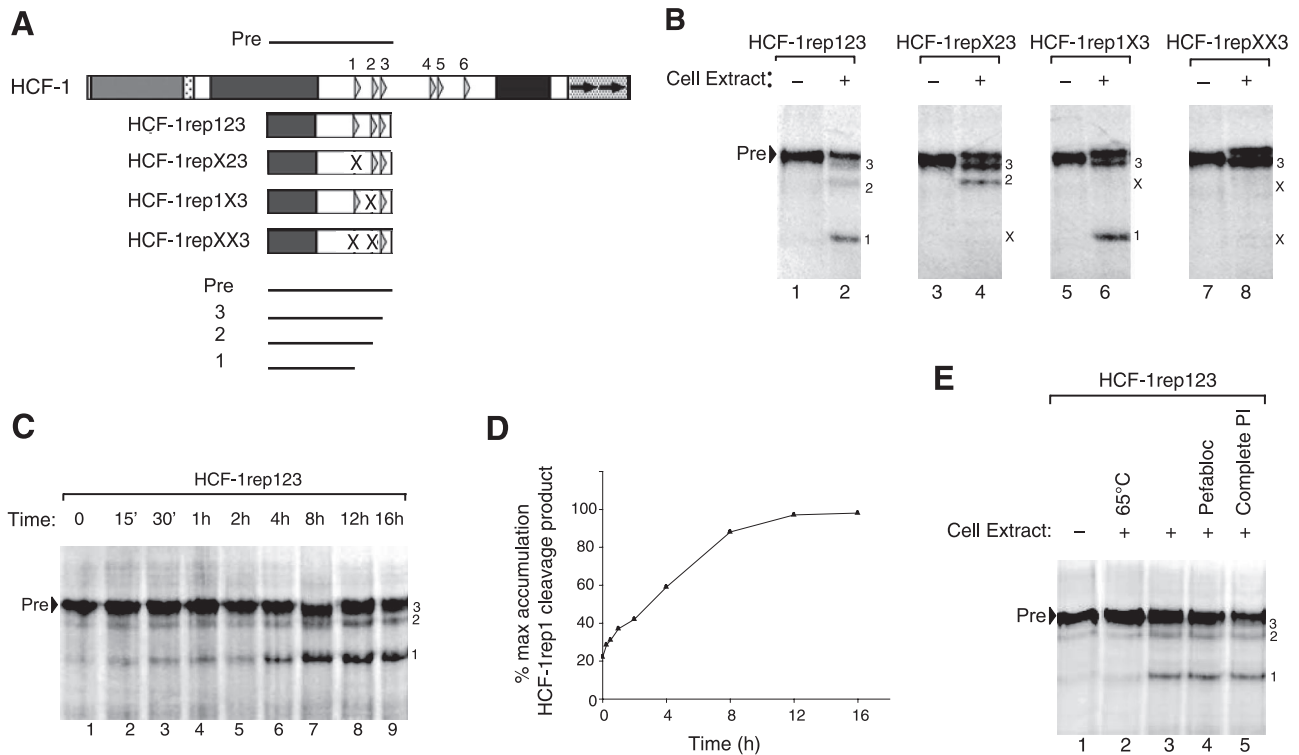


FIG. 5. HCF-1<sub>PRO</sub> repeats are faithfully processed by human cell extracts. (A) Schematic of HCF-1 cleavage precursors. The line labeled Pre indicates the region contained within each precursor. These HCF-1 precursors contain the first three HCF-1<sub>PRO</sub> repeats in a wild-type version (HCF-1rep123) or with mutated repeat 1 (HCF-1repX23), mutated repeat 2 (HCF-1rep1X3), or mutated repeats 1 and 2 (HCF-1repXX3). Products corresponding to full-length precursor (Pre) and N-terminal cleavage products for HCF-1<sub>PRO</sub> repeats 3, 2, and 1 are shown below. (B) HeLa cell extract cleavage of HCF-1<sub>PRO</sub> repeats. <sup>35</sup>S-labeled HCF-1 precursors from panel A were incubated without (lanes 1, 3, 5, and 7) or with (lanes 2, 4, 6, and 8) nuclear HeLa cell extract. Numbers on the right of each panel indicate the N-terminal cleavage product derived from the indicated HCF-1<sub>PRO</sub> repeat; x indicates the missing cleavage product corresponding to the mutated HCF-1<sub>PRO</sub> repeat; C-terminal fragments are not visible owing to their reduced specific activity. (C) Time course of the HeLa cell extract proteolytic activity. <sup>35</sup>S-labeled HCF-1rep123 precursor was incubated with nuclear HeLa cell extracts for the indicated periods of time. (D) Quantification of time course data. The relative accumulation of HCF-1<sub>PRO</sub> repeat 1 cleavage product from panel C was quantified to represent HCF-1<sub>PRO</sub> repeat cleavage activity over time. (E) Characterization of the HeLa cell proteolytic activity. <sup>35</sup>S-labeled HCF-1rep123 precursor was incubated without extract (lane 1) or with nuclear HeLa cell extract (lanes 2 to 5) that was heat treated (lane 2), untreated (lane 3), or treated with Pefabloc (lane 4) or Complete protease inhibitor (PI) mixture (lane 5).

trithorax group proteins human MLL and *Drosophila* Trx and the HCF proteins human HCF-1 and dHCF. We have shown that, like MLL, the *Drosophila* Trx is a substrate for the taspase 1 protease. Interestingly, the enzyme-substrate specificity, i.e., human taspase 1 with MLL and *Drosophila* taspase 1 with Trx, has coevolved. In contrast, with respect to HCF proteins, very different proteolytic pathways have evolved. Thus, the dHCF protein is cleaved by *Drosophila* taspase 1 but its human (and *Xenopus*) HCF-1 counterparts are cleaved by what is likely an unusual proteolytic activity.

**Evolution of taspase 1 enzyme-substrate specificity.** Hsieh et al. (10) noted that sequences encoding the taspase 1 enzymes have been conserved in vertebrates and insects but not in the worm *Caenorhabditis elegans* and that, correspondingly, the MLL/Trx developmental regulators in vertebrates and insects but not in worms have taspase 1 recognition sites. Consistent with this observation, Trx is indeed a taspase 1 substrate at its CS2-like taspase 1 cleavage site. Examination of the activities of the human and *Drosophila* taspase 1 enzymes on the MLL and Trx proteins also revealed a more refined coevolution of enzyme and substrate, as human taspase 1 is more active on its

cognate MLL substrate and *Drosophila* taspase 1 is more active on its cognate Trx substrate. This observation suggests that the presence of the taspase 1 enzyme and MLL/Trx protein maturation have been conserved because the cleavage is critically important for proper development (16, 19, 26). Nevertheless, there is considerable flexibility in the interaction between protease and substrate, and for such enzyme-substrate coevolutionary flexibility to be possible, it is likely that taspase 1 does not possess a large number of essential targets.

We note with interest that vertebrates and insects display common longitudinal body segmentation, which is lacking in *C. elegans*, and that proper MLL and Trx maturation is functionally linked to proper homeodomain gene expression, which is intimately involved in segmentation determination (10, 16, 19, 26). Consistent with the suggestion that MLL/Trx cleavage is required to control the expression of genes involved in body segmentation, in the sea urchin *Strongylocentrotus purpuratus*, which, like *C. elegans*, is not longitudinally segmented, the Trx homolog lacks an evident taspase 1 recognition site and there is no evident taspase 1-encoding gene (data not shown).

Although HCF proteins are not known to be involved in

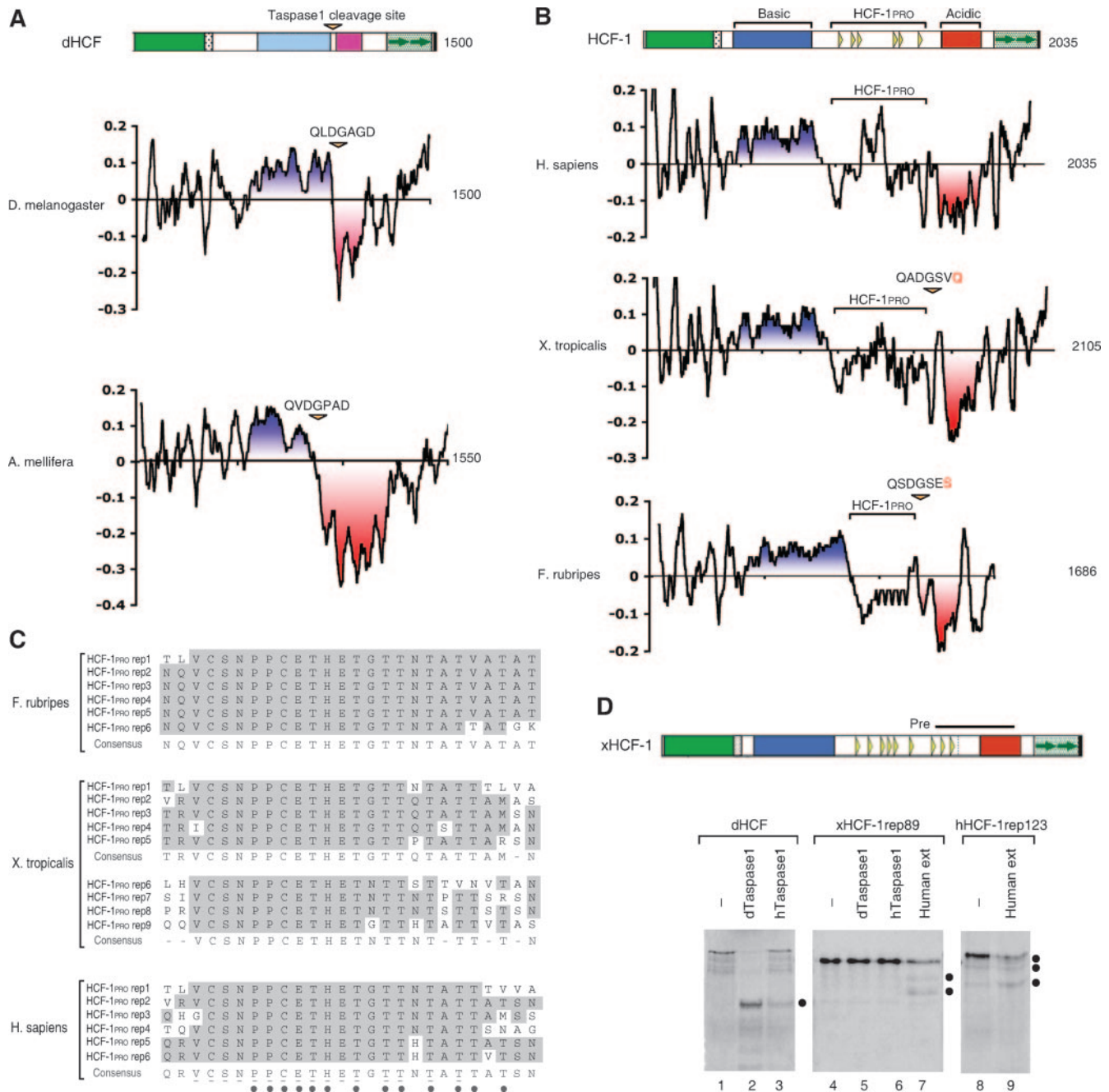


FIG. 6. Evolution of HCF protein maturation. (A) HCF protein maturation in insects. Top, schematic representation of dHCF. The taspase 1 cleavage site is indicated by the arrowhead. Bottom, charge profiles of the fly *Drosophila melanogaster* and the honeybee *Apis mellifera* HCF proteins. Peaks above zero indicate basic regions, and peaks below zero indicate acidic regions; basic and acidic regions are shown in blue and red, respectively. For each protein, sequences of taspase 1 cleavage sites are indicated above arrowheads. (B) HCF protein maturation in vertebrates. Top, schematic representation of human HCF-1. Segments corresponding to the basic and acidic regions and to the HCF-1<sub>PRO</sub> repeats are indicated above the schematic. Bottom, charge profiles of the human *Homo sapiens*, frog *Xenopus tropicalis*, and fish *Fugus rubripes* HCF proteins as in panel A. For each protein, the region corresponding to the HCF-1<sub>PRO</sub> repeats is overlined. Sequences corresponding to partially conserved taspase 1 cleavage sites are indicated above arrowheads; in red is indicated the residue that does not match the taspase 1 site consensus. (C) Sequence conservation of the HCF-1<sub>PRO</sub> repeats in vertebrates. An alignment of fish (top), frog (center), and human (bottom) HCF-1<sub>PRO</sub> repeats is shown. The human HCF-1<sub>PRO</sub> repeats are numbered as in reference 15. Residues matching a consensus based on the most frequent residue at each position are shaded. Positions in the consensus sequences for which a conserved residue cannot be defined are indicated by dashes. Underlined positions indicate sequence conservation among HCF-1<sub>PRO</sub> repeats of all three species. ●, residues important for human HCF-1<sub>PRO</sub> repeat cleavage in vivo (32). (D) In vitro cleavage of xHCF-1. Top, schematic representation of xHCF-1. The line labeled Pre indicates the region contained in the cleavage precursor. The dashed line indicates the imperfect taspase 1 cleavage site. Bottom, xHCF-1 precursor (xHCF-1rep89) was incubated without additions (lane 4) or with either *Drosophila* taspase 1 (dTaspase1) (lane 5), human taspase 1 (hTaspase1) (lane 6), or HeLa cell extract (lane 7). dHCF (lanes 1 to 3) and human HCF-1 (hHCF-1rep123; lanes 8 and 9) precursors are shown as positive controls for taspase 1 (dHCF) and HCF-1<sub>PRO</sub> repeat (hHCF-1rep123) cleavage. ●, HCF protein cleavage products.

regulating segmentation, invertebrate HCF proteins display species-specific patterns of proteolytic maturation that follow the Trx proteins. Thus, in *Drosophila*, dHCF and Trx are both cleaved by *Drosophila* taspase 1, and in the honeybee the HCF and Trx homologs possess taspase 1 recognition sites. In contrast, as with the Trx homologs, the sea urchin and worm HCF proteins do not possess evident taspase 1 recognition sites (data not shown); indeed, worm HCF does not undergo proteolytic maturation (34). We suggest that in invertebrates, the HCF and Trx proteins coevolved taspase 1-induced maturation. In contrast, in vertebrates, HCF-1 proteins took on a very different proteolytic maturation mechanism. Nevertheless, the parallels between MLL/Trx and HCF protein maturation suggest that in addition to regulation of the cell cycle, HCF-1 may also play a role in regulation of genes involved in segment determination.

**HCF-1 is likely cleaved by an unusual protease.** We have described an enzymatic activity in HeLa cell extracts that specifically cleaves the HCF-1<sub>PRO</sub> repeat. We do not know at this time the precise relationship between the HeLa cell activity described here and the HCF-1 autocatalytic activity described previously (28), but we note some differences between the two activities. The autocatalytic activity, which depends on sequences within the HCF-1<sub>C</sub> region, results in cleavage at not only HCF-1<sub>PRO</sub> repeats but also an additional site, called 102, which lacks evident HCF-1<sub>PRO</sub> repeat sequence similarity and is located just C terminal of the sixth HCF-1<sub>PRO</sub> repeat (15, 28). This autocatalytic site is, however, not used in vivo when the HCF-1<sub>PRO</sub> repeats are inactivated by mutation or deletion (14, 29, 32), and we have not observed 102 site cleavage using a precursor substrate spanning this site in the HeLa cell extract assay (F. Capotosti and W. Herr, unpublished results). Thus, in both HeLa cells and HeLa cell extracts, HCF-1 proteolytic maturation appears to be HCF-1<sub>PRO</sub> repeat specific. We also note that, unlike the autocatalytic 102 site activity, which is sensitive to the serine protease inhibitor Pefabloc (28), the HeLa cell HCF-1<sub>PRO</sub> repeat activity is resistant to this protease inhibitor (Fig. 5E). Clarifying the relationship between the HeLa cell activity described here and the HCF-1 autocatalytic activity will require further investigation.

Whatever the relationship, however, the properties of the HCF-1<sub>PRO</sub> repeat protease are likely to be unusual. As shown in Fig. 6C, there are 16 amino acid residues over 19 positions that are identical among the HCF-1<sub>PRO</sub> repeat consensus sequence of the fish, frog, and human HCF-1 proteins, an impressive level of sequence conservation for a proteolytic recognition sequence. Furthermore, as described previously (32), alanine substitutions at 12 positions over 18 of the human HCF-1<sub>PRO</sub> repeat 2 affect HCF-1<sub>PRO</sub> repeat cleavage in HeLa cells (Fig. 6C), a surprisingly large sequence requirement for proteolytic cleavage. Given this extensive sequence conservation and sequence requirements, we suggest that the functional HCF-1<sub>PRO</sub> repeat element reflects the assemblage of different recognition sequences for (i) the protease itself and (ii) accessory factors that could regulate proteolysis. Such accessory factors could directly recruit or stimulate the protease or could reflect enzymes that modify the HCF-1<sub>PRO</sub> repeat (e.g., by phosphorylation or glycosylation) to regulate cleavage. Whatever the reason, the large size of the HCF-1<sub>PRO</sub> repeat results in considerable specificity, as we have been unable to find a

match to the HCF-1<sub>PRO</sub> repeat in any other protein sequence found in public protein databases, suggesting that in toto the mechanisms of HCF-1<sub>PRO</sub> repeat cleavage are unique to HCF-1 maturation.

**Alternate mechanisms of HCF protein cleavage result in HCF<sub>N</sub> and HCF<sub>C</sub> subunits of similar structure.** The evolutionary comparisons shown in Fig. 6 indicate that although the mechanisms of HCF protein maturation have switched between insects and vertebrates, the end results of proteolytic processing are similar. To illustrate this point, Fig. 6A and B show charge profiles for the two insect (fly and honeybee) and three vertebrate (fish, frog, and human) HCF proteins. These profiles show that the insect and vertebrate HCF proteins all contain basic and acidic regions at corresponding positions. This conservation of amino acid composition is consistent with the important cellular functions these regions have been shown to possess in human HCF-1: G<sub>1</sub> phase progression in the case of the basic region (30) and M phase progression (13), transcriptional activation (17), and chromatin association (13) in the case of the acidic region. We note with interest that whether the HCF protein is cleaved by taspase 1, as appears to be the case with the insect HCF proteins, or at the HCF-1<sub>PRO</sub> repeats, as appears to be the case in the vertebrate proteins, the cleavage site(s) is always positioned between the basic and acidic regions. Thus, although the mechanism for HCF protein proteolysis has changed during evolution, the resulting HCF<sub>N</sub> and HCF<sub>C</sub> subunits are very similar in structure.

**How and why might a transition from taspase 1- to HCF-1<sub>PRO</sub> repeat-dependent HCF-1 proteolysis have evolved?** The unexpected switch in proteolytic processing mechanism between insect and vertebrate HCF proteins leads to the questions of how and why. Concerning how the HCF-1<sub>PRO</sub> repeats may have arisen, we note with interest that the six HCF-1<sub>PRO</sub> repeats in fish and human HCF-1 are all encoded by a single large exon of 1,477 bp in human. (The nine repeats in *X. tropicalis* are encoded by two exons [five in the first and four in the second], which may have resulted from a duplication of the six-repeat-containing fish/human exon [Fig. 6C].) We imagine that an HCF-1 progenitor acquired the six HCF-1<sub>PRO</sub> repeats as a single genetic element by recombination, perhaps transposition, prior to vertebrate evolutionary divergence.

Concerning why a switch may have occurred, one possible explanation is that the region containing the HCF-1<sub>PRO</sub> repeats has been evolutionarily selected as an additional platform for protein-protein interactions, as Vogel and Kristie (29) have shown that the transcriptional coactivator/corepressor FHL2 interacts with nonprocessed HCF-1, stimulating transcription of an HCF-1 target gene. This interaction between HCF-1<sub>PRO</sub> repeats and cofactors could allow the modulation of HCF-1 processing and activity. Another possible explanation is that, with an HCF-1<sub>PRO</sub> repeat protease, HCF-1 protein maturation has become independent from taspase 1 and thus from MLL/Trx maturation. This could result in a more flexible regulation of these two important cell cycle regulators. In any case, however, the importance of HCF-1<sub>PRO</sub> repeat processing during vertebrate evolution is underscored by the remarkable similarity among the repeats themselves in one species (e.g., human) and between vertebrates as divergent as fish and human (Fig. 6C).

In conclusion, in human and fly, the MLL/Trx and HCF



proteins have conserved the process of proteolytic maturation, but the MLL/Trx process coevolved with the cognate taspase 1 proteases, whereas HCF-1 and dHCF have apparently evolved very different proteolytic pathways. Whatever the reason for the evolutionary change to HCF-1<sub>PRO</sub> repeat-dependent processing in vertebrates, it is evident that the HCF-1<sub>PRO</sub> repeats are unusual cleavage sites, and it is likely that the protease responsible for their cleavage has unusual properties.

#### ACKNOWLEDGMENTS

We thank Angus Wilson for the dHCF expression vector; Deborah Auferio for help with characterizing the anti-dHCF antisera; Nouria Hernandez, Christina Hertel, Sara Rodriguez Jato, and Shweta Tyagi for critical readings of the manuscript; and Henrik Kaessmann for advice.

F.C. is the recipient of a Roche Research Foundation Fellowship. These studies were supported by the Swiss National Science Foundation and the University of Lausanne.

#### REFERENCES

- Ayton, P. M., and M. L. Cleary. 2001. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene* **20**:5695–5707.
- Breen, T. R., and P. J. Harte. 1991. Molecular characterization of the trithorax gene, a positive regulator of homeotic gene expression in *Drosophila*. *Mech. Dev.* **35**:113–127.
- Brown, M. S., and J. L. Goldstein. 1997. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**:331–340.
- Dignam, J. D., R. M. Lebovitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**:1475–1489.
- Djabali, M., L. Selleri, P. Parry, M. Bower, B. D. Young, and G. A. Evans. 1992. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat. Genet.* **2**:113–118.
- Goto, H., S. Motomura, A. C. Wilson, R. N. Freiman, Y. Nakabeppu, K. Fukushima, M. Fujishima, W. Herr, and T. Nishimoto. 1997. A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. *Genes Dev.* **11**:726–737.
- Goulet, B., and A. Nepveu. 2004. Complete and limited proteolysis in cell cycle progression. *Cell Cycle* **3**:986–989.
- Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C. M. Croce, and E. Canaani. 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.
- Guelman, S., T. Suganuma, L. Florens, S. K. Swanson, C. L. Kiesecker, T. Kusch, S. Anderson, J. R. Yates III, M. P. Washburn, S. M. Abmayr, and J. L. Workman. 2006. Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. *Mol. Cell. Biol.* **26**:871–882.
- Hsieh, J. J., E. H. Cheng, and S. J. Korsmeyer. 2003. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**:293–303.
- Hsieh, J. J., P. Ernst, H. Erdjument-Bromage, P. Tempst, and S. J. Korsmeyer. 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.
- Izeta, A., S. Malcomber, and P. O'Hare. 2003. Primary structure and compartmentalization of *Drosophila melanogaster* host cell factor. *Gene* **305**:175–183.
- Julien, E., and W. Herr. 2004. A switch in mitotic histone H4 lysine 20 methylation status is linked to M phase defects upon loss of HCF-1. *Mol. Cell* **14**:713–725.
- Julien, E., and W. Herr. 2003. Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *EMBO J.* **22**:2360–2369.
- Kristie, T. M., J. L. Pomerantz, T. C. Twomey, S. A. Parent, and P. A. Sharp. 1995. The cellular C1 factor of the herpes simplex virus enhancer complex is a family of polypeptides. *J. Biol. Chem.* **270**:4387–4394.
- Kuzin, B., S. Tillib, Y. Sedkov, L. Mizrokhi, and A. Mazo. 1994. The *Drosophila* trithorax gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene fork head. *Genes Dev.* **8**:2478–2490.
- Luciano, R. L., and A. C. Wilson. 2002. An activation domain in the C-terminal subunit of HCF-1 is important for transactivation by VP16 and LZIP. *Proc. Natl. Acad. Sci. USA* **99**:13403–13408.
- Mahajan, S. S., K. M. Johnson, and A. C. Wilson. 2003. Molecular cloning of *Drosophila* HCF reveals proteolytic processing and self-association of the encoded protein. *J. Cell. Physiol.* **194**:117–126.
- Mazo, A. M., D. H. Huang, B. A. Mozer, and I. B. Dawid. 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., S. D. Briggs, H. W. Brock, M. E. Martin, D. Gibbs, C. D. Allis, and J. L. Hess. 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell* **10**:1107–1117.
- Nakamura, T., T. Mori, S. Tada, W. Krajewski, T. Rozovskaia, R. Wassell, G. Dubois, A. Mazo, C. M. Croce, and E. Canaani. 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell* **10**:1119–1128.
- Patel, T., G. J. Gores, and S. H. Kaufmann. 1996. The role of proteases during apoptosis. *FASEB J.* **10**:587–597.
- Reilly, P. T., and W. Herr. 2002. Spontaneous reversion of tsBN67 cell proliferation and cytokinesis defects in the absence of HCF-1 function. *Exp. Cell Res.* **277**:119–130.
- Schroeter, E. H., J. A. Kisslinger, and R. Kopan. 1998. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**:382–386.
- Sedkov, Y., S. Tillib, L. Mizrokhi, and A. Mazo. 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.
- Takeda, S., D. Y. Chen, T. D. Westergard, J. K. Fisher, J. A. Rubens, S. Sasagawa, J. T. Kan, S. J. Korsmeyer, E. H. Cheng, and J. J. Hsieh. 2006. Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes Dev.* **20**:2397–2409.
- Tyagi, S., A. L. Chabes, J. Wysocka, and W. Herr. 2007. E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol. Cell* **27**:107–119.
- Vogel, J. L., and T. M. Kristie. 2000. Autocatalytic proteolysis of the transcription factor-coactivator C1 (HCF): a potential role for proteolytic regulation of coactivator function. *Proc. Natl. Acad. Sci. USA* **97**:9425–9430.
- Vogel, J. L., and T. M. Kristie. 2006. Site-specific proteolysis of the transcriptional coactivator HCF-1 can regulate its interaction with protein cofactors. *Proc. Natl. Acad. Sci. USA* **103**:6817–6822.
- Wilson, A. C., R. N. Freiman, H. Goto, T. Nishimoto, and W. Herr. 1997. VP16 targets an amino-terminal domain of HCF involved in cell cycle progression. *Mol. Cell. Biol.* **17**:6139–6146.
- Wilson, A. C., K. LaMarco, M. G. Peterson, and W. Herr. 1993. The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74**:115–125.
- Wilson, A. C., M. G. Peterson, and W. Herr. 1995. The HCF repeat is an unusual proteolytic cleavage signal. *Genes Dev.* **9**:2445–2458.
- Wysocka, J., and W. Herr. 2003. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem. Sci.* **28**:294–304.
- Wysocka, J., Y. Liu, R. Kobayashi, and W. Herr. 2001. Developmental and cell-cycle regulation of *Caenorhabditis elegans* HCF phosphorylation. *Biochemistry* **40**:5786–5794.
- Yokoyama, A., I. Kitabayashi, P. M. Ayton, M. L. Cleary, and M. Ohki. 2002. Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. *Blood* **100**:3710–3718.
- Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D. J. Auferio, I. Kitabayashi, W. Herr, and M. L. Cleary. 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell. Biol.* **24**:5639–5649.
- Yu, B. D., J. L. Hess, S. E. Horning, G. A. Brown, and S. J. Korsmeyer. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**:505–508.
- Zhou, H., S. Spicuglia, J. J. Hsieh, D. J. Mitsiou, T. Hoiby, G. J. Veenstra, S. J. Korsmeyer, and H. G. Stunnenberg. 2006. Uncleaved TFIID is a substrate for taspase 1 and active in transcription. *Mol. Cell. Biol.* **26**:2728–2735.