Evolutionary emergence of Hairless as a novel component of the Notch signaling pathway

Steven W Miller¹, Artem Movsesyan¹, Sui Zhang¹, Rosa Fernández²,³, James W Posakony¹

¹Division of Biological Sciences, Section of Cell and Developmental Biology, University of California San Diego, La Jolla, United States
²Bioinformatics and Genomics Unit, Center for Genomic Regulation, Barcelona, Spain
³Current address: Department of Life Sciences, Barcelona Supercomputing Center, Barcelona, Spain

For correspondence: jposakony@ucsd.edu
Abstract

Suppressor of Hairless [Su(H)], the transcription factor at the end of the Notch pathway in *Drosophila*, utilizes the Hairless protein to recruit two co-repressors, Groucho (Gro) and C-terminal Binding Protein (CtBP), indirectly. Hairless is present only in the Pancrustacea, raising the question of how Su(H) in other protostomes gains repressive function. We show that Su(H) from a wide array of arthropods, molluscs, and annelids includes motifs that directly bind Gro and CtBP; thus, direct co-repressor recruitment is ancestral in the protostomes. How did Hairless come to replace this ancestral paradigm? Our discovery of a protein (S-CAP) in Myriapods and Chelicerates that contains a motif similar to the Su(H)-binding domain in Hairless has revealed a likely evolutionary connection between Hairless and Metastasis-associated (MTA) protein, a component of the NuRD complex. Sequence comparison and widely conserved microsynteny suggest that S-CAP and Hairless arose from a tandem duplication of an ancestral MTA gene.
Introduction

A very common paradigm in the regulation of animal development is that DNA-binding transcriptional repressors bear defined amino acid sequence motifs that permit them to recruit, by direct interaction, one or more common co-repressor proteins that are responsible for conferring repressive activity. Two such universal co-repressors are Groucho (Gro) and C-terminal Binding Protein (CtBP).

The ancient and highly conserved transcription factor Suppressor of Hairless [Su(H)] functions at the terminus of the widely utilized Notch cell-cell signaling pathway. Su(H) is converted into an activator by signaling through the Notch receptor, but in the absence of signaling it functions as a repressor. Earlier studies have revealed that in many settings in Drosophila, Su(H)’s repressive activity depends on binding to the Hairless protein (Figure 1). Hairless includes separate Gro- and CtBP-binding motifs, which permit it to function as an adaptor to bring these two corepressors to Su(H) (Figure 1B) (Barolo et al. 2002). Thus, the Su(H)/H partnership in the fly represents a notable exception to the rule of direct co-repressor recruitment.

As genome and transcriptome sequences have become available for more and more insects and other arthropods, we have searched for possible Hairless orthologs in a wide variety of species, in an attempt to determine the protein’s phylogenetic distribution. We have found that Hairless is confined to the Pancrustacea (or Tetraconata), a clade of arthropods that includes the Crustacea and Hexapoda (Misof et al. 2014; Kjer et al. 2016). While this indicates that Hairless was gained at least 500 Mya,
it also raises the question of how Su(H) in other protostomes acquires repressive activity.

Here we present evidence that direct co-repressor recruitment by Su(H) is likely to be ancestral in the protostomes. We show that Su(H) in a broad range of protostomes, including arthropods, molluscs, and annelids, bears both a short linear motif that mediates binding of CtBP and a novel motif for direct recruitment of Gro. Thus, the evolutionary appearance of Hairless has permitted the replacement of an ancient and predominant regulatory mechanism (direct co-repressor recruitment) with a novel one (indirect recruitment).

What can we learn about the evolutionary history of Hairless? While Hairless itself is found only in the Pancrustacea, we show that the genomes of Myriapods and Chelicerates encode a protein with clear sequence and functional similarities to Hairless. These proteins include a motif that strongly resembles the Su(H)-binding domain of Hairless, and we demonstrate that this motif from the house spider *Parasteatoda tepidariorum* does indeed bind Su(H). In addition, these Myriapod and Chelicerate proteins also include one or more canonical motifs for recruitment of CtBP. Accordingly, we designate these factors as “Su(H)-Co-repressor Adaptor Proteins” (S-CAPs).

Finally, further sequence analyses, along with the discovery of conserved microsynteny, have provided substantial evidence that Hairless and the S-CAPs are likely to be homologous and that they arose from a duplication of the gene encoding
Metastasis-associated (MTA) protein, a component of the nucleosome remodeling and deacetylase (NuRD) complex.

An intriguing question in evolutionary biology concerns the path by which a particular clade has escaped a strongly selected character that has been conserved for hundreds of millions of years. We believe that our study has yielded valuable insight into both the emergence of an evolutionary novelty and its replacement of an ancestral paradigm.

Results

Hairless is present only in the Pancrustacea

We have conducted extensive BLAST searches of genome and transcriptome sequence data for a wide variety of metazoa in an attempt to define the phylogenetic distribution of Hairless. We find that Hairless as originally described (Bang and Posakony 1992; Maier et al. 1992; Maier et al. 2008) is confined to the Pancrustacea (or Tetraconata), and occurs widely within this clade, including the Hexapoda, Vericrustacea, and Oligostraca (Figure 2A). By contrast, no evidence for a true Hairless gene has been detected in either Myriapods or Chelicerates, even in cases where substantially complete genome sequence assemblies are available.

The enormous variation in the size of the Hairless protein in various Pancrustacean clades is worthy of note (Figure 1A). The known extremes are represented by the Diplostracan (shrimp) Eulimnadia texana (343 aa) (Baldwin-Brown et al. 2018) and the Dipteran (fly) Protophormia terraenovae (1614 aa) (Hase et al. 2017), a 4.7-fold difference.
There is a broad tendency for the size of the protein to be relatively stable within an order (Supplementary file 1). Thus, as noted previously (Maier et al. 2008), the Hymenoptera generally have a small Hairless (of the order of 400 aa; see Figure 1A), while the Diptera typically have a much larger version (of the order of 1000 aa or more). Notable exceptions to this pattern of uniformity are aphids, where Hairless is typically ~900 aa compared to ~400 aa in other Hemiptera, and chalcid wasps, where the protein is over 500 aa instead of the Hymenoptera-typical ~400 aa noted above (Supplementary file 1). Smaller Hairless proteins retain all five conserved motifs/domains characteristic of this factor (Maier et al. 2008), while the regions that flank and lie between these sequences are reduced in size (Figure 1A; Supplementary file 2).

A known CtBP-binding motif is present in the non-conserved N-terminal region of Su(H) in a wide variety of protostomes

The apparent confinement of the Hairless co-repressor adaptor protein to the Pancrustacea raises the question of the mechanism(s) by which Su(H) in other protostomes might recruit co-repressor proteins to mediate its repressor function. Of course, other protostomes need not utilize the Gro and CtBP co-repressors for this purpose; different co-repressors might substitute. Nevertheless, we first sought to identify known binding motifs for Gro and CtBP in Su(H) from arthropods lacking Hairless. As shown in Table 1, we found a canonical CtBP recruitment motif of the form PφDφS (where φ = I, L, M, or V) in predicted Su(H) proteins from a variety of Myriapods and Chelicerates, including the centipede Strigamia maritima, the tick Ixodes scapularis, the spider Parasteatoda tepidariorum, the horseshoe crab Limulus polyphemus,
and the scorpion *Centruroides sculpturatus*. These motifs are all located in the non-conserved N-terminal region of Su(H) (Supplementary file 3).

Extending this sequence analysis to other protostome phyla led to the finding that a similar PψDψS motif occurs in the N-terminal region of Su(H) from a large number of molluscs and annelids, as well as from multiple Nemertea, Brachiopoda, Phoronida, and monogonont rotifers, and also from some flatworms (Table 1). It is notable, by contrast, that we do not find CtBP-binding motifs present in Su(H) from nematodes. Nevertheless, given the broad phylogenetic distribution of the PψDψS motif in Su(H) from both Ecdysozoa and Lophotrochozoa, our observations strongly suggest that direct recruitment of CtBP by Su(H) is ancestral in the protostomes.

To verify that the shared PψDψS motif in protostome Su(H) proteins can indeed mediate direct recruitment of CtBP, we carried out an in vitro pulldown assay using GST-tagged *Drosophila* CtBP (bound to Glutathione Sepharose beads) and a His-tagged fragment of *Strigamia maritima* Su(H) (Figure 3A). We found that the two proteins do interact directly and robustly, in a manner that is dependent on the integrity of the PVDLS motif in *Strigamia* Su(H).

**A novel conserved motif in protostome Su(H) binds the Gro co-repressor**

In addition to a PψDψS CtBP-binding motif, we have found that Su(H) from a wide variety of protostomes includes a novel motif similar to GSLTPPDKV (Table 1). Where present, this sequence typically lies a short (but variable) distance C-terminal to the PψDψS motif, also within the non-conserved N-terminal region of the protein.
The GSLTPPKV motif is particularly prevalent in Su(H) from the Trochozoa, which includes annelids, sipunculans, molluscs, nemerteans, brachiopods, and phorons (Kocot et al. 2017). Among the Ecdysozoa, it appears consistently in Su(H) from Crustacea and Myriapoda, and in small subsets of both Hexapoda (Ephemeroptera, Odonata, Zygentoma, Archaeognatha, Diplura, and Collembola) and Chelicerata [harvestmen (Opiliones) and Scorpiones]. The motif is absent from Su(H) in all other insect orders, and we have not found it so far in Su(H) from nematodes, flatworms, rotifers, or tardigrades; it is, however, found in the onychophoran Euperipatoides kanangrensis (Table 1). Perhaps surprisingly, the motif is present in Su(H) from the acorn worms Saccoglossus kowalevskii and Ptychodera flava (Simakov et al. 2015), which are hemichordates (deuterostomes).

Using an in vitro pulldown assay, we tested the possibility that the GSLTPPKV motif mediates binding of the Gro co-repressor (Figure 3B). Indeed, we find that GST-tagged Gro protein interacts strongly with a His-tagged protein bearing this motif at its C-terminus, and that this binding is abolished when the motif is replaced by alanine residues. We conclude that Su(H) from a broad range of protostomes is capable of directly recruiting both CtBP and Gro (Table 1), and that this capacity is hence very likely to be ancestral in this clade.

**Retention of the hybrid state: Species that have both Hairless and the co-repressor-binding motifs in Su(H)**

The evolutionary emergence of Hairless as an adaptor protein capable of mediating the indirect recruitment of both Gro and CtBP to Su(H) might be expected to relieve a
selective pressure to retain the ancestral Gro- and CtBP-binding motifs in Su(H) itself.

And indeed, we find that Su(H) from multiple insect orders comprising the Neoptera lacks both of these sequences (Figure 2B). Strikingly, however, we have observed that Crustacea and a small group of Hexapoda retain both traits (Figure 2B). Thus, multiple representatives of the Branchiopoda, Malacostraca, and Copepoda, along with Ephemeroptera, Odonata, Zygoptera, Archaeognatha, Diplura, and Collembola, have both a canonical Hairless protein (including its Gro- and CtBP-binding motifs) and Gro- and CtBP-binding motifs within Su(H). These clades, then, appear to have retained a "hybrid intermediate" state (Baker et al. 2012) characterized by the presence of both corepressor recruitment mechanisms.

Myriapods and Chelicerates encode a protein with similarity to Hairless

While canonical Hairless proteins are confined to the Pancrustacea, we have discovered that the genomes of Myriapods and Chelicerates nevertheless encode a protein with intriguing similarities to Hairless. Most notable is the presence of a motif that strongly resembles the "Su(H) binding domain" (SBD) of Hairless, which mediates its high-affinity direct interaction with Su(H) (Figure 1; Figure 4A). We will refer to these proteins as "S-CAPs"; the basis for this designation will be made clear in forthcoming figures. We note that the occurrence of this protein in the centipede Strigamia maritima has also recently been reported by Maier (Maier 2019). In the Pancrustacea, the N-terminal and C-terminal halves of the Hairless SBD are encoded by separate exons (Figure 4B). Strikingly, the related motif in Myriapod and Chelicerate S-CAPs is
likewise encoded by separate exons, with exactly the same splice junction as in Hairless (Figure 4B). We believe that this is highly unlikely to be coincidental, and is instead strongly suggestive of an evolutionary relationship between Hairless and S-CAPs.

A recent structural analysis of the Su(H)-Hairless protein complex identified several residues in the Hairless SBD that are involved in binding to the C-terminal domain (CTD) of Su(H) (Yuan et al. 2016) (Figure 4A). These include four hydrophobic amino acids in the main body of the SBD (L235, F237, L245, and L247; these are highlighted in red in Figure 4A). Note that the Myriapod and Chelicerate S-CAP motifs share these same residues. In addition, a tryptophan (W258) C-terminal to the main body of the Hairless SBD also participates in binding to Su(H) (Figure 4A). Myriapod and Chelicerate S-CAPs all include a tryptophan residue at a similar position C-terminal to the main SBD-like domain (Figure 4A). Moreover, this particular W residue in both Hairless and the S-CAPs is followed by a hydrophobic residue, typically V or I. These sequence features, we suggest, is further strong evidence of a common ancestry for the respective segments of Hairless and S-CAPs.

A third structural similarity between Hairless and S-CAPs is the presence in the latter of one or more short linear motifs capable of binding the CtBP co-repressor (Figure 5A). These motifs typically reside in the C-terminal half of the S-CAPs, superficially resembling the C-terminal location of Hairless's CtBP recruitment motif.

A table listing representative examples of Myriapod and Chelicerate S-CAPs is provided as Supplementary file 4, and an annotated FASTA file of their amino acid
Sequences is included as Supplementary file 5. It is important to note that we have not found non-Hairless S-CAPs in the Pancrustacea.

Spider S-CAP binds to *Drosophila* Su(H)

Given the clear sequence similarity between the Hairless SBD and the SBD-like motif in Myriapod and Chelicerate S-CAPs, we investigated whether the latter motif is likewise capable of mediating direct binding to Su(H). As noted above, the Hairless SBD interacts specifically with the CTD of Su(H). Since this domain in Su(H) is very highly conserved throughout the Bilateria and Cnidaria, we thought it reasonable to utilize *Drosophila* Su(H) for this binding assay. As shown in Figure 4C, we find that a 200-amino-acid segment of S-CAP from the spider *Parasteatoda tepidariorum* binds directly to *Drosophila* Su(H) in vitro. This interaction depends strictly on the integrity of the five residues that in Hairless have been shown to contact the Su(H) CTD (highlighted in red in Figure 4A).

Given the presence of one or more CtBP recruitment motifs in the Myriapod and Chelicerate S-CAP proteins (Figure 5A), along with the ability of their SBD-like domains to bind Su(H) (Figure 4C), we have designated these as “Su(H)-Co-repressor Adaptor Proteins” (S-CAPs).

Chelicerate S-CAP proteins are related to Metastasis-Associated (MTA) proteins

In addition to their similarities to Hairless, the S-CAP proteins of Chelicerates include two regions with strong sequence homology to the Metastasis-associated (MTA) protein family, which is highly conserved among Metazoa. The MTA proteins play an
important role in transcriptional regulation via their function as core components of the nucleosome remodeling and deacetylase (NuRD) complex (Allen et al. 2013; Burgold et al. 2019). The N-terminal half of MTAs includes four well-defined functional domains: BAH (Bromo-Adjacent Homology), ELM2 (Egl-27 and MTA1 homology), SANT (Swi3, Ada2, N-CoR, and TFIIB), and GATA-like zinc finger (Millard et al. 2014) (Figure 5B). Of these, the ELM2 and SANT domains are retained at the N-terminal end of Chelicerate S-CAPs (Figure 5B; Figure S1A). This is highly likely to have functional significance, as the ELM2 and SANT domains of MTA proteins work together to recruit and activate the histone deacetylases HDAC1 and HDAC2 (Millard et al. 2013). Further suggesting homology between Chelicerate S-CAPs and MTAs is the observation that their shared ELM2 and SANT domains are each encoded by two exons with exactly the same splice junction (Figure 5C).

It is noteworthy that, despite sharing the SBD-like and CtBP recruitment motifs of Chelicerate S-CAPs, the available Myriapod S-CAP protein sequences lack the N-terminal ELM2 and SANT homologies with MTA proteins (Figure 5B). Consistent with this, the SBD motif in Myriapod S-CAPs lies much closer to the protein’s N terminus than the SBD motif in Chelicerate S-CAPs, suggesting that simple loss of the ELM2/SANT-encoding exons might underlie this difference between the two S-CAP clades. Likewise, Hairless proteins are devoid of clear similarities to MTAs.

In addition to their SBD and ELM2/SANT domains, Chelicerate S-CAPs share a third region of homology that lies between the ELM2 and SANT sequences (Figure 5-figure supplement 1A). This region is absent from both Hairless and the Myriapod S-
CAPs. Conversely, Myriapod S-CAPs include three segments of sequence similarity in their C-terminal regions that are not found in either Hairless or Chelicerate S-CAPs (Figure 5-figure supplement 1B).

Conserved microsynteny between MTA and S-CAP/Hairless genes

Our analysis of the genomic locations of genes encoding MTA proteins in Arthropoda, Hairless in Pancrustacea, and S-CAPs in Myriapods and Chelicerates has yielded the surprising finding that proximate or near-proximate linkage between MTA and Hairless genes or between MTA and S-CAP genes is broadly conserved among arthropods (Figure 6; Supplementary file 1; Supplementary file 4). Thus, in the centipede Strigamia maritima, the gene encoding S-CAP lies immediately upstream of that encoding MTA, in the same orientation (Figure 6; Supplementary file 4). A similar linkage relationship between S-CAP and MTA genes is seen in many arachnids, including the spiders Nephila clavipes (Supplementary file 4) and Parasteatoda tepidariorum (Figure 6; Supplementary file 4) and the mites Achipteria coleoptrata and Sarcoptes scabiei (Supplementary file 4). Likely due at least in part to its history of whole-genome duplication (Nossa et al. 2014; Kenny et al. 2016), the horseshoe crab Limulus polyphemus (representing the Merostomata/Xiphosura) has three paralogous copies of this same S-CAP-MTA linkage pairing (Supplementary file 4). Some exceptions to this pattern do exist. In the genomes of the mites Metaseiulus occidentalis (Supplementary file 4) and Varroa destructor (Techer et al. 2019), for example, the genes encoding S-CAP and MTA are far separated from each other.
Close, typically adjacent, linkage between *Hairless* and *MTA* genes is likewise widely observed in the genomes of Pancrustacea. Among the Hexapoda, this pattern can be found in many different orders (Supplementary file 1), including Diptera, Lepidoptera, Coleoptera (Figure 6), Hymenoptera (Figure 6), Psocodea, Hemiptera (Figure 6), Thysanoptera, Blattodea, Orthoptera, Odonata, and Collembola. Among the Vericrustacea, adjacent linkage of *Hairless* and *MTA* is seen in the shrimp *Triops cancriformis* (Notostraca) (Supplementary file 1). Nevertheless, exceptions are readily found, even within the same orders as above (Supplementary file 1). Examples include *Drosophila melanogaster, Ceratitis capitata, and Lucilia cuprina* (Diptera; Supplementary file 1), *Bicyclus anynana* (Lepidoptera), *Anoplophora glabripennis, Dendroctonus ponderosae*, and *Nicrophorus vespilloides* (Coleoptera), and *Cimex lectularius* (Hemiptera; Supplementary file 1).

Interestingly, in some instances *Hairless/MTA* microsynteny is preserved, but the genes’ relative orientation is different (Figure 6; Supplementary file 1). Thus, in the aphids — in contrast to other Hemiptera — *MTA* lies downstream of *Hairless*, but in the opposite orientation (Figure 6). In the beetle *Harmonia axyridis* (Coleoptera), *MTA* lies upstream of *Hairless* (Figure 6).

Despite the multiple instances in which it has been lost, we believe that the most parsimonious interpretation of our analysis is that close linkage between *MTA* and *S-CAP/Hairless* genes is ancestral in the respective taxa (Myriapods/Chelicerates and Pancrustacea). We leave for the Discussion our proposed interpretation of the evolutionary significance of this adjacency.
Discussion

The evolution of Hairless represents a shift from the ancestral and dominant paradigm of direct co-repressor recruitment by Su(H)

Our analysis of sequences from a broad range of protostomes strongly suggests that direct recruitment of the CtBP and Gro co-repressors by Su(H) is ancestral in this clade. This is consonant with the fact that direct co-repressor recruitment by DNA-binding repressor proteins in general is a dominant paradigm among Metazoa. This evokes the intriguing question of what might have led to the loss of direct recruitment by Su(H) in the Neoptera (see Figure 1B) and its replacement by Hairless-mediated indirect recruitment? Does Hairless provide some advantageous functional capacity? Note that this is not intended to suggest that Hairless must be an evolutionary adaptation per se (Lynch 2007); rather, we are asking: What capability might it have conferred that would lead to its retention and the subsequent loss of the recruitment motifs in Su(H)?

One appealing (but of course speculative) possibility is that Hairless may have permitted Su(H) for the first time to recruit both CtBP and Gro simultaneously to the same target genes. As we have noted, the apparently ancestral P\$D\$S and GSLTPDPKV motifs in protostome Su(H) typically lie quite close to each other in the protein’s linear sequence (Supplementary file 3). CtBP (~400 aa) and Gro (~700 aa) are both large proteins that engage in oligomerization as part of their functional mechanism (Song et al. 2004; Bhambhani et al. 2011). It is very unlikely that both could bind stably to DNA-bound Su(H) at the same time. In contrast, the Gro and CtBP recruitment motifs in Hairless are far apart in the linear sequence (Figure 1A) and are separated by a
region predicted to be largely disordered (Figure 1-figure supplement 1). We suggest that this might be compatible with simultaneous recruitment of the two co-repressors. Whatever may have been the selective forces that led to the loss of direct co-repressor recruitment by Su(H) in the Neoptera and its replacement by Hairless-mediated indirect recruitment, Hairless is a notable evolutionary novelty for having permitted the unusual abandonment of an ancestral and highly conserved paradigm. We suggest that this represents a striking example of “developmental system drift” (True and Haag 2001), in which a common output (widespread “default repression” of Notch pathway target genes) is achieved via distinct molecular mechanisms in different species.

A possible evolutionary pathway for the appearance of Hairless

We have described here several findings that we believe have important implications for an attempt to reconstruct the history of Hairless as an evolutionary novelty. First, we observe that Hairless is apparently confined to the Pancrustacea, wherein it is widely distributed among diverse taxa (Figure 2A; Supplementary file 1). Second, we have discovered in the sister groups Myriapoda and Chelicerata a protein (S-CAP) with clear sequence homology to the Su(H) binding domain (SBD) of Hairless (Figure 4A). Significantly, in both Hairless and the S-CAPs these motifs are encoded by contributions from two exons, with the associated splice junction in precisely the same location (Figure 4B; Supplementary file 4). Third, we find that S-CAPs in the Chelicerata include in their N-terminal region strong homology to the ELM2 and SANT
domains of MTAs, which themselves are highly conserved among Metazoa, and therefore would have been present in the arthropod common ancestor (Figure 5B,C).

Finally, our analysis indicates that close, usually adjacent, linkage of *Hairless* and *MTA* genes (in the Pancrustacea) and between *S-CAP* and *MTA* genes (in the Myriapoda and Chelicerata) is widespread (Figure 6; Supplementary file 1; Supplementary file 4), and hence very likely to be ancestral, in these taxa.

While any attempt to infer the sequence of evolutionary events that led to the appearance of *Hairless* is necessarily speculative, we believe that the above findings offer substantial support for the following hypothetical pathway. We propose that in a deep arthropod ancestor a tandem duplication of the *MTA* gene occurred. One copy retained the strong sequence conservation (and presumably ancestral function) of metazoan *MTA* genes, while the second copy diverged very substantially, eventually encoding a protein that had lost all but the ELM2 and SANT domains of the MTA ancestor. The extensive reconfiguration of this paralog also included the eventual acquisition of the SBD motif and the addition of one or more CtBP recruitment motifs (see Figure 7 for some possible sources of these components). In the Myriapod lineage, even the ELM2 and SANT domains were eventually lost. In the Pancrustacea, we suggest that this same divergent *MTA* paralog evolved to become *Hairless*. Beyond the alterations described for the Myriapoda, this would have involved the acquisition of sequences encoding additional now-conserved domains and motifs, including the Gro recruitment motif (Supplementary file 2). This radical evolutionary transformation
resulted in a protein with little or no remaining homology to its MTA ancestor, and
with an entirely novel regulatory function (Holland et al. 2017).

In this context, it is of interest that the Drosophila Mi-2/Nurd complex — which includes the MTA protein — has recently been shown to engage in direct repression of multiple Notch pathway target genes, independent of both Su(H) and Hairless (Zacharioudaki et al. 2019). Whether this activity preceded the emergence of Hairless is unknown, but the possibility that it is in some way connected to Hairless’s evolutionary history is indeed intriguing.

**Materials and methods**

<table>
<thead>
<tr>
<th>Key Resources Table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reagent type (species) or resource</strong></td>
</tr>
<tr>
<td>Antibody</td>
</tr>
<tr>
<td>Antibody</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>HIS-HLHmBetaSmarMUT pRSET-C clone</td>
</tr>
<tr>
<td>HIS-HLHmBetaWT pRSET-C clone</td>
</tr>
<tr>
<td>HIS-HLHmBetatrunc pRSET-C clone</td>
</tr>
<tr>
<td>HIS-PtepSCAP233-432 WT pRSET-C clone</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
<tr>
<td>Recombinant DNA reagent</td>
</tr>
</tbody>
</table>

Sigma (formerly Amersham; discontinued)
<table>
<thead>
<tr>
<th>Recombinant DNA reagent</th>
<th>pRSET-C</th>
<th>Invitrogen</th>
<th>CAT#V35120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial assay or kit</td>
<td>Chem Illumination Reagents</td>
<td>Pierce ECL Western Blotting Substrate</td>
<td>CAT#32209</td>
</tr>
<tr>
<td>Resource, sequence database</td>
<td>NCBI</td>
<td>NCBI</td>
<td>RRID:SCR_006472</td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>NCBI BLAST</td>
<td>NCBI</td>
<td>RRID:SCR_004870</td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>DNA Strider</td>
<td>DOI: 10.1093/nar/16.5.1829; DOI: 10.1007/BF02821333</td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>BlastStation-Local64</td>
<td>TM Software, Inc.</td>
<td></td>
</tr>
</tbody>
</table>
Sequence searches, analysis, and annotation

Genome and transcriptome sequences encoding Hairless, Suppressor of Hairless, S-CAP, and MTA proteins from a wide variety of species were recovered via BLAST searches, using either the online version at the NCBI website (Boratyn et al. 2013) or the version implemented by the BlastStation-Local64 desktop application (TM Software, Inc.). Sequences were analyzed and annotated using the GenePalette (Rebeiz and Posakony 2004; Smith et al. 2017) and DNA Strider (Marck 1988; Douglas 1995) desktop software tools. Analysis of predicted disordered regions in Hairless was conducted using DISOPRED3 on the PSIPRED server (Buchan et al. 2013; Jones and Cozzetto 2015).

Generation of constructs for GST pulldown experiments

Strigamia maritima Su(H) protein constructs to test CtBP binding

A codon-optimized fragment corresponding to exons 2 and 3 from S. maritima Su(H) mRNA was synthesized by Genewiz, Inc., and cloned into pRSET-C using Acc65I and BamHI restriction sites. The CtBP-motif mutant was subsequently generated by overlap extension PCR (Ho) using the primers HISsmarSUH-f (CGCTGGATTCGCCGACATGAC), HISsmarSUH-r (CCATGGTACCAGTTATGCGTGG), HISsmarSUHctbpm-f (AACCACgCCGcaGcTGcGgCTAACAGCCATCGCGGTGAAGGCGGCCAC), HISsmarSUHctbpm-r (GCTGTTAGcGcCGcTGcCGTGGTGTTGTCGCGAAGTGAAGCGGTCAG). After sequence confirmation, this fragment was also cloned into pRSET-C using the same
enzymes. Binding of these constructs to *Drosophila melanogaster* CtBP was assayed using GST alone and a GST-CtBP fusion protein (Nibu et al. 1998).

**Constructs to test potential Gro-binding motif in *Strigamia maritima* Su(H)**

A truncated version of HLHm$\beta$ (HLHm$\beta$-WRPWtrunc) was amplified from a pRSET-HLHm$\beta$-WT construct using the primers HISmbeta-f (cgatggatccgaATGGTTCTGGAAATGGAGATGTCCAAG) and HISmbetatrunc-r (ccatggtaccagTCACATGGGGCCagagttgagctgccgctgggccgc); a version of HLHm$\beta$ with the WRPW motif replaced with the amino acids GSLTPPDKV (HLHm$\beta$+Smar-motifWT) was amplified from the WT construct with HISmbeta-f and mbetaSmarSuH-r (ccatggatccgaTCACACTTTATCAGGTGAAGAGAGAAACCATGGGGCCagagttggagctgccg); and a version of HLHm$\beta$ with the WRPW motif replace with a stretch of 9 alanine residues (HLHm$\beta$+Smar-motifMUT) was amplified using HISmbeta-f and mbetaSmarSuHmut-r (ccatggatccgaTCAGgtgccgtgccgctgccgtgcATGGGGCCagagttggagctgcc). Each construct was then subsequently cloned into pRSET-C using the restriction enzymes BamHI and Acc65I and sequence verified. Binding of these constructs to *Drosophila melanogaster* Gro was assayed using GST alone and a GST-Gro fusion protein. The latter construct was made by cloning the full-length Gro coding sequence into the pGEX-KG expression vector at the EcoRI and SalI restriction sites:

gtggcgaccatccatccaaatcggtctcgttcgctggatccccgggaatttcccggtggtggtggtggaattcATG...

TAAATCCACAAAAACATGCAGTTTTTTTCATTTGTAAAAAGCTCGTATAGTTTTT
S-CAP/Hairless constructs for Su(H) interaction analysis

Codon-optimized fragments from *Drosophila melanogaster* Hairless (residues 192-389), and *Parasteatoda tepidariorum* cS-CAP (residues 233-432) as well as 5-alanine mutant substitutions (Dmel: GGLRQFFKDGBKFILELARSKDGDKSWG -> GGRAQAFKDGBKFILEARSKDGDKSGA; Ptep: VGSLKFLFLGRLVLKLNAAQDDGSGNKCQW -> VGSKAFLGGRLVAKANAAQDDGSGNKCQA) were synthesized by Genewiz, Inc.

Inserts were subsequently cloned into pRSET-C using the restriction enzymes BamHI and Acc65I. Binding of these constructs to *Drosophila melanogaster* Su(H) was assayed using GST alone and a GST-Su(H) fusion protein (Bailey and Posakony 1995).

GST pulldowns using each of these constructs were performed as previously described (Fontana and Posakony 2009).

Synthesized, codon-optimized sequences

>Smar Su(H)ex2-3 WT (116 aa)

CGCTGGATCCGCGGCCAGTATGACTACCCGCCGCCGTTAGCCAGCAACATACACGCGAGGCCGA
CCTGTGGAACGTGAACCTGGCCACCTACAGCAGCGCACCGACCACATGCACCGGTGCAACCCCGG
CACCTAGCGTTACCGGTTTCTACGCCCAGGCCACCGGCAGCAACAGCGTTAGCCCGAGTAGCGTG
AGCCTGACCACCCTGACCCCTCACTTCGCCGACAACCACCGTGGACCTGAGCAACAGCCATCG
CGGTGAAGGCGGCCACCTGGATCTGGTGCGCTTCCAGAGCGACCGCGTGGATGCCTACAAGCAG
CCAACGGCCTGAGCGTGCATATCCCGGACCACCACGCACTGGACCTGAGCAACCGCATCG
Smar Su(H)ex2-3 mut

>CGCTGGATACCGCGCCGCGCACTGGACATACCCCGCCGCCGCTGATTCCAGCAGACATACGCGCGAGGCCGA

CCTGTGGAACGTGAACCTGGCCACCTACAGCAGCGCACCGACCACA

TGCACCGGTGCAACCCCGG

CACCTAGCGTTACCGGTTTCTACGCCCAGGCCACCGGCAGCAACAGCGTTAGCCCGAGTAGCGTG

AGCCTGACCACCCTGACCCCTCACTTCGCCGACAACCACgCCGcaGcTGcGgCTAACAGCCATCG

CGGTGAAGGCGGCCACCTGGATCTGGTGCGCTTCCAGAGCGACCGCGTGGATGCCTACAAGCACG

CCAACCGCTAGGCGTGCATATCCCGGACCACACGCTAATCTGGTACCATTG

Dmel Hairless192-389 WT

CGATGGATCCGAGCAGTGGTTGCAGCAGCAGCTGGCACTGCCAAAATTGGTAAAGGCAGCAACAG

CGGTGGCAGTTTTGATATGGGCCGCACACCGATCAGCACCCATGGCAACAACAGCTGGGGTGGCT

ATGGCGGCCGTTTACAGTTCTTTAAAGATGGCAAGTTTATTTTAGAACTGGCCCGCAGCAAAGAT

GGCGATAAAAGCGGCTGGGTGAG

TGTGACCCGCAAAACCTTTCGCCGCCGAGTGCAGCAACCAG

CGCAACCGTGACCCCTACCAGTGCCGTGACCACCGCCTACCCGAAGAATGAAAACAGCACCTCTT

TAAGCTTTAGCGACGACAATAGCAGCATTCAGAGCAGCCCGTGGCAGCGCGATCAGCCGTGGAAA

CAGAGTCGTCCGCGCCGTGGCATCAGCAAAGAACTGTCTTTATTTTTCCACCGCCCGCAATAG

TACAACGGGTGTCGTGACGCTTTAGCTACCGGACGCCGCAAAAGCTGTGCTGCCCGCATGAAACCGCTGA

CCACCAGCGAAGATCAGACGCCGATCTTTTGCCAACCGCAATCAAAGCCGAGACGGTGTGATGACT

TTAAAAAGCGCAAGACGCCGAATAACTGTTACCATTG

Dmel Hairless192-389 5Amut

CGATGGATCCGAGCAGCAGCAGCTGGCACTGCCAAAATTGGTAAAGGCAGCAATAG

CGGTGGCAGTTTTGACATGGGCCGCACACCGATCAGCACCCATGGCAACAACAGCTGGGGTGGCT

ATGGCGGCCGTTTACAGTTCTTTAAAGACGGCAAGTTCATCGCCGAAGCCGCACGCAGCAAAGAT

GGCGACAAAAGCGGCTGGGTGAGCGTGACCCGCAAAACCTTTCGCCGCCGAGTGCAGCAACCAG

CGCAACCGTGACCCCTACCAGTGCCGTGACCACCGCCTACCCGAAAAACGAAAACAGCACCTCTT

TAAGCTTTTAGCGACGACAATAGCAGCATTCAGAGCAGCCCGTGGCAGCGCGATCAGCCGTGGAAA

CAGAGTCGTCCGCGCCGTGGCATCAGCAAAGAACTGTCTTTATTTTTCCACCGCCCGCAATAG

TACAACGGGTGTCGTGACGCTTTAGCTACCGGACGCCGCAAAAGCTGTGCTGCCCGCATGAAACCGCTGA

CCACCAGCGAAGATCAGACGCCGATCTTTTGCCAACCGCAATCAAAGCCGAGACGGTGTGATGACT

TTAAAAAGCGCAAGACGCCGAATAACTGTTACCATTG

Ptep S-CAP233-432 WT

CGATGGATCCGAGCAGCAGCAGCTGGCACTGCCAAAATTGGTAAAGGCAGCAATAG

CGGTGGCAGTTTTGATATGGGCCGCACACCGATCAGCACCCATGGCAACAACAGCTGGGGTGGCT

ATGGCGGCCGTTTACAGTTCTTTAAAGATGGCAAGTTTATTTTAGAACTGGCCCGCAGCAAAGAT

GGCGATAAAAGCGGCTGGGTGAGCGTGACCCGCAAAACCTTTCGCCGCCGAGTGCAGCAACCAG

CGCAACCGTGACCCCTACCAGTGCCGTGACCACCGCCTACCCGAAAAACGAAAACAGCACCTCTT

TAAGCTTTTAGCGACGACAATAGCAGCATTCAGAGCAGCCCGTGGCAGCGCGATCAGCCGTGGAAA

CAGAGTCGTCCGCGCCGTGGCATCAGCAAAGAACTGTCTTTATTTTTCCACCGCCCGCAATAG

TACAACGGGTGTCGTGACGCTTTAGCTACCGGACGCCGCAAAAGCTGTGCTGCCCGCATGAAACCGCTGA

CCACCAGCGAAGATCAGACGCCGATCTTTTGCCAACCGCAATCAAAGCCGAGACGGTGTGATGACT

TTAAAAAGCGCAAGACGCCGAATAACTGTTACCATTG

Ptep S-CAP233-432 5Amut

CGATGGATCCGAGCAGCAGCAGCTGGCACTGCCAAAATTGGTAAAGGCAGCAATAG

CGGTGGCAGTTTTGACATGGGCCGCACACCGATCAGCACCCATGGCAACAACAGCTGGGGTGGCT

ATGGCGGCCGTTTACAGTTCTTTAAAGACGGCAAGTTCATCGCCGAAGCCGCACGCAGCAAAGAT

GGCGACAAAAGCGGCTGGGTGAGCGTGACCCGCAAAACCTTTCGCCGCCGAGTGCAGCAACCAG

CGCAACCGTGACCCCTACCAGTGCCGTGACCACCGCCTACCCGAAAAACGAAAACAGCACCTCTT

TAAGCTTTTAGCGACGACAATAGCAGCATTCAGAGCAGCCCGTGGCAGCGCGATCAGCCGTGGAAA

CAGAGTCGTCCGCGCCGTGGCATCAGCAAAGAACTGTCTTTATTTTTCCACCGCCCGCAATAG

TACAACGGGTGTCGTGACGCTTTAGCTACCGGACGCCGCAAAAGCTGTGCTGCCCGCATGAAACCGCTGA

CCACCAGCGAAGATCAGACGCCGATCTTTTGCCAACCGCAATCAAAGCCGAGACGGTGTGATGACT

TTAAAAAGCGCAAGACGCCGAATAACTGTTACCATTG
ACCGCTGAAAAAAGGCATGACACCAGCAGTGGCAGATCCGAGCGCATTAAGAAA
CCCGTTTAAAGGAGTACGAGACCAGCAGAAAACAGTGCTTTAGTTTACTGCTGTGCAGCAGCAG
TTGGACACCGCCGGTAGCTGCAAGAAAGTATCGATGTGGACGACACCAGCAGCAAACCA
GCGAAGGCTACATCAGCCCCGATTCTGAGCAGCAATAGCCGACCAGCAAATTTGATACCATTTAA
CATGATTTTGCAAGCAATCCGATACCTAATGTCGATACCATGG

513
514
515
516
517
518
519
Acknowledgements

We are especially grateful to our many colleagues who made their genome and transcriptome sequence assemblies freely available; without their generosity, this study would not have been possible. We thank Scott Rifkin and Mark Rebeiz for helpful discussion and input during the preparation of the manuscript. We also thank the following artists for making available the illustrations shown in Figure 2A: crab, by Firkin (https://openclipart.org/detail/270221/crab-silhouette); monarch butterfly, by carolemagnet (https://openclipart.org/detail/263384/monarch-butterfly); centipede, by Firkin (https://openclipart.org/detail/261126/centipede-3); spider, by liftarn (https://openclipart.org/detail/179190/spider); horseshoe crab, by Gosc (https://openclipart.org/detail/174556/horseshoe-crab); tick, by Juhele (https://openclipart.org/detail/279073/simple-tick-ixodes-ricinus-silhouette). This work was supported by NIH Grants R01GM046993 and R01GM120377 (to JWP).
References


Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV et al. 2010. Genome sequences of the...


<table>
<thead>
<tr>
<th>Species</th>
<th>CtBP motif</th>
<th>Gro motif</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ecdyonurus insignis</strong></td>
<td>YPDNH[PVDLS]SPRPH</td>
<td>APMI[PGL]TPPD[MK]NGE[H]PHHG</td>
<td>1</td>
</tr>
<tr>
<td><strong>Calopteryx splendens</strong></td>
<td>YTDNH[PVDLS]SPPRPP</td>
<td>HHMIPGSLTPPD[MK]NGE[H]PMH</td>
<td>2</td>
</tr>
<tr>
<td><strong>Atelura formicaria</strong></td>
<td>YPDNH[PVDLS]SPRPF</td>
<td>PHMIPGSLTPPD[MK]NGE[H]HSS</td>
<td>3</td>
</tr>
<tr>
<td><strong>Machilis hrabei</strong></td>
<td>YPDNH[PVDLS]SPRPH</td>
<td>PHMIPGSLTPPD[MK]NGE[H]PHHG</td>
<td>4</td>
</tr>
<tr>
<td><strong>Catajayps aquilonaris</strong></td>
<td>STANN[PVDLS]SPRGS</td>
<td>APMIPGSLTPPD[K]VNGE[H]HSSH</td>
<td>5</td>
</tr>
<tr>
<td><strong>Holacanthella duospinosa</strong></td>
<td>VPNSN[PVDLS]NPSPS</td>
<td>SNFVGSLPPPERKM[NG]MPS</td>
<td>6</td>
</tr>
<tr>
<td><strong>Hyalella azteca</strong></td>
<td>SLGHR[PVDLS]SNSH</td>
<td>YSHSLTPPDVDGNSMQ</td>
<td>8</td>
</tr>
<tr>
<td><strong>Eurytemora affinis</strong></td>
<td>SETSA[PVDLS]SNARS</td>
<td>YGMLPGSLTPPDKLNGHCSPS</td>
<td>9</td>
</tr>
<tr>
<td><strong>Triops cancricornis</strong></td>
<td>HPEAR[PVDLS]SSL</td>
<td>YSS[TL]TPPDVDGNSMQ</td>
<td>10</td>
</tr>
<tr>
<td><strong>Centruroides sculpturatus</strong></td>
<td>GYESSPVDLS[SH]RVS</td>
<td>MQLI[GS]MS[TS]HDKVG[ND][H]SLS</td>
<td>19</td>
</tr>
<tr>
<td><strong>Euperipatoides kanangrensis</strong></td>
<td>NSYN[PVDLS]SRRRL</td>
<td>Q[Q]ILPGSLPPDKVNGD[VLSLA]</td>
<td>20</td>
</tr>
<tr>
<td><strong>Naisneris dendritica</strong></td>
<td>DPNGHPVDLS[HRS]HI</td>
<td>PHMIPGSLTPPD[DRVK]NGE[PSSGL]</td>
<td>21</td>
</tr>
<tr>
<td><strong>Brachionus koreanus</strong></td>
<td>AKELETPIDLS[SK]S</td>
<td>[V]VSMA</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 1. Co-repressor recruitment motifs in protostome Su(H) proteins

Shown are alignments of short linear amino acid motifs (bold) in the N-terminal region of Su(H) proteins that mediate direct recruitment of the co-repressors CtBP and Gro.

Numbers in the column at right refer to the source of the corresponding sequence data, with accession numbers and publication citations indicated (see list below).

1. GCCL01029953.1 (Simon et al. 2018)
2. LYUA01002621.1 (Ioannidis et al. 2017)
3. GAYJ02050375.1 (Misof et al. 2014)
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1058</td>
<td>4</td>
<td>Scaffold 1 (i5K 2013)</td>
</tr>
<tr>
<td>1059</td>
<td>5</td>
<td>JYFJ02000853.1 (i5K 2013)</td>
</tr>
<tr>
<td>1060</td>
<td>6</td>
<td>NIPM01000059.1 (Wu et al. 2017a)</td>
</tr>
<tr>
<td>1061</td>
<td>7</td>
<td>GGJN01104381.1 (unpublished)</td>
</tr>
<tr>
<td>1062</td>
<td>8</td>
<td>NW_017238139.1 (i5K 2013)</td>
</tr>
<tr>
<td>1063</td>
<td>9</td>
<td>NW_019396104.1 (i5K 2013)</td>
</tr>
<tr>
<td>1064</td>
<td>10</td>
<td>BAYF01001879.1 (Ikeda et al. 2015)</td>
</tr>
<tr>
<td>1065</td>
<td>11</td>
<td>JW959185.1 (Sahoo et al. 2013)</td>
</tr>
<tr>
<td>1066</td>
<td>12</td>
<td>JH430541.1 (Chipman et al. 2014)</td>
</tr>
<tr>
<td>1067</td>
<td>13</td>
<td>NW_003805473.1 (Hoy et al. 2016)</td>
</tr>
<tr>
<td>1068</td>
<td>14</td>
<td>NW_002722632.1 (Gulia-Nuss et al. 2016)</td>
</tr>
<tr>
<td>1069</td>
<td>15</td>
<td>NW_018383625.1 (Schwager et al. 2017)</td>
</tr>
<tr>
<td>1070</td>
<td>16</td>
<td>NW_018370942.1 (Schwager et al. 2017)</td>
</tr>
<tr>
<td>1071</td>
<td>17</td>
<td>NW_013671976.1 (Battelle et al. 2016)</td>
</tr>
<tr>
<td>1072</td>
<td>18</td>
<td>NW_013676581.1 (Battelle et al. 2016)</td>
</tr>
<tr>
<td>1073</td>
<td>19</td>
<td>NW_019384406.1 (Schwager et al. 2017)</td>
</tr>
<tr>
<td>1074</td>
<td>20</td>
<td>LN881712.1 (unpublished)</td>
</tr>
<tr>
<td>1075</td>
<td>21</td>
<td>(Andrade et al. 2015)</td>
</tr>
<tr>
<td>1076</td>
<td>22</td>
<td>KP293861.1 (Gazave et al. 2017)</td>
</tr>
<tr>
<td>1077</td>
<td>23</td>
<td>(Lemer et al. 2015)</td>
</tr>
<tr>
<td>1078</td>
<td>24</td>
<td>(Andrade et al. 2015)</td>
</tr>
<tr>
<td>1079</td>
<td>25</td>
<td>NW_014678436.1 (Albertin et al. 2015)</td>
</tr>
<tr>
<td>1080</td>
<td>26</td>
<td>NW_011936122.1 (Zhang et al. 2012)</td>
</tr>
<tr>
<td>1081</td>
<td>27</td>
<td>NW_008708401.1 (Simakov et al. 2013)</td>
</tr>
<tr>
<td>1082</td>
<td>28</td>
<td>GFRY01035878.1 (Luo et al. 2018)</td>
</tr>
<tr>
<td>1083</td>
<td>29</td>
<td>(Whelan et al. 2014)</td>
</tr>
<tr>
<td>1084</td>
<td>30</td>
<td>GDJY01029776.1 (Luo et al. 2015)</td>
</tr>
<tr>
<td>1085</td>
<td>31</td>
<td>GFSC01078935.1 (Luo et al. 2018)</td>
</tr>
<tr>
<td>1086</td>
<td>32</td>
<td>GAKZ01044347.1 (unpublished)</td>
</tr>
<tr>
<td>1087</td>
<td>33</td>
<td>GBXV02009219.1 (Lee et al. 2015)</td>
</tr>
<tr>
<td>1088</td>
<td>34</td>
<td>(Brauchle et al. 2018)</td>
</tr>
</tbody>
</table>
**Figure legends**

**Figure 1.** Hairless mediates indirect recruitment of co-repressor proteins to Su(H). (A) Diagram denoting locations of conserved domains and motifs within Hairless, and illustrating extreme size differences of the protein in different species. Shown is Hairless from the carpenter bee *Ceratina calcarata* and the blowfly *Protophormia terraenovae* (Hase et al. 2017), with scale and protein sizes indicated. (B) Summary of Hairless’s known mode of action (Lai 2002; Maier 2006) as an adaptor protein that recruits the global co-repressors C-terminal Binding Protein (CtBP) and Groucho (Gro) to Suppressor of Hairless [Su(H)], the transducing transcription factor for the Notch (N) cell-cell signaling pathway; adapted from Figure 6 of (Barolo et al. 2002). In the absence of signaling through the Notch receptor (left), Su(H) acts as a repressor of Notch target genes, despite the presence of transcriptional activator proteins (orange oval). Upon activation of the Notch receptor (middle), Su(H), in a complex with the receptor’s intracellular domain (NICD) and the co-activator Mastermind (Mam), functions to activate transcription of pathway target genes in cooperation with other transcriptional activators. In the absence of Hairless and hence in the absence of Su(H)’s repressive activity (right), the partner transcription factors are often sufficient to activate expression of target genes in a signal-independent manner (Barolo and Posakony 2002).

**Figure 1-figure supplement 1.** Graph showing predicted disordered regions in *Drosophila melanogaster* Hairless, generated by DISOPRED3 (Buchan et al. 2013; Jones and Cozzetto 2015). Locations of the Gro-binding motif (purple arrow) and CtBP-
binding motif (red arrow) are indicated. Note that most of the region between these
two motifs is strongly predicted to be disordered.

**Figure 2.** Phylogenetic distribution of Hairless and related S-CAP proteins. (A) Based
on extensive BLAST searches of available genome and transcriptome assemblies,
orthologs of the canonical Hairless gene are found only in the Pancrustacea (blue bar),
while orthologs of a gene that encodes the related S-CAP protein are found in the
Myriapods (mS-CAP, red bar) and Chelicerates (cS-CAP, pink bar). We suggest S-CAP
as a suitable umbrella nomenclature for this gene family (black bracket). Tree adapted
from Figure 2 of Regier et al. (Regier et al. 2010). (B) Consistent with the presence of
Hairless as an adaptor protein, Su(H) in most insect orders (the Neoptera clade) has lost
the ancestral short linear motifs that mediate direct recruitment of the CtBP and Gro co-
repressor proteins (red bar). However, in the Crustacea, Collembola, Diplura, and a
subset of Insecta, the ancestral recruitment motifs have been retained in Su(H), despite
the presence of Hairless (see Table 1 and Supplementary file 3). Tree adapted from
Misof et al. (Misof et al. 2014) and Kjer et al. (Kjer et al. 2016).

**Figure 3.** Direct binding of co-repressor proteins by Su(H) from the centipede Strigamia
maritima. (A) The PVDLS motif in the N-terminal region of Su(H) from the centipede
Strigamia maritima directly binds Drosophila CtBP. A His-tagged 116-aa segment of the
Strigamia Su(H) protein, bearing a PVDLS recruitment motif for CtBP, binds strongly to
GST-dCtBP (WT, lane 2). Mutation of the motif to alanines (AAAAA) abolishes this
interaction (PVDLSmut, lane 1). The results shown in this panel have been replicated in eight additional experiments, utilizing three independent isolations of GST-CtBP protein from bacterial cultures. **(B)** The conserved GSLTPPDKV motif in the N-terminal region of *Strigamia* Su(H) directly binds *Drosophila* Gro. His-tagged E(spl)mβ-HLH protein, which bears a C-terminal WRPW motif that recruits Gro, is used as a binding control. Wild-type (WT) HLHmβ binds GST-Gro (lane 2), while a truncated version of the protein lacking the WRPW motif (lane 1) fails to bind. A synthetic version of HLHmβ in which the WRPW motif has been replaced by the wild-type GSLTPPDKV motif also binds GST-Gro efficiently (lane 4), while a mutant version in which GSLTPPDKV is replaced by alanines (AAAAAAAAA) shows extremely weak binding (lane 3). No binding of any of the His-tagged proteins to GST alone is observed, even with substantially greater amounts of GST compared to GST-Gro. The results shown in this panel have been replicated in seven additional experiments, utilizing five independent isolations of GST-Gro protein from bacterial cultures.

**Figure 4.** S-CAP proteins in Myriapods and Chelicerates contain a Hairless-like domain that binds Su(H). **(A)** Alignment of the Suppressor of Hairless Binding Domain (SBD) in *Drosophila melanogaster* (Dmel) Hairless with the related motif in the S-CAP proteins from a representative set of Myriapods and Chelicerates. Numbers flanking each sequence segment represent amino acid positions within the protein. The contiguous SBD motif is highlighted in bold. Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs.
Amino acids in Hairless that have been shown to make direct contact with Su(H) [including the non-contiguous tryptophan (W) residue] (Yuan et al. 2016) are highlighted in red. Hydrophobic residues nearly always found immediately adjacent to the W are underlined. Species names are as follows: Smar (Strigamia maritima); Hsub (Hydroschendyla submarina) (Fernández et al. 2016); Agir (Anopsobius giribeti) (Fernández et al. 2016); Ptep (Parasteatoda tepidariorum); Hruf (Hypochthonius rufulus) (Bast et al. 2016); Ssca (Sarcoptes scabiei); Vjac (Varroa jacobsoni) (Techer et al. 2019); Iper (Ixodes persulcatus); Mmar (Mesobuthus martensii) (Cao et al. 2013); Lpol (Limulus polyphemus).

We note that Maier (Maier 2019) has previously described the presence of the SBD-like element in the Strigamia maritima sequence. (B) SBD motifs in both Hairless and S-CAP proteins (red) are encoded in two exons with the same splice junction (indicated by /; see grey highlight). Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs. Species names as in A, except for Isca (Ixodes scapularis). (C) Spider S-CAP protein binds directly to Drosophila Su(H) in vitro. In all panels, lanes 1-4 represent the indicated His-tagged segments of wild-type Drosophila (Dmel) Hairless (lane 1); Dmel Hairless bearing alanine substitutions for each of five SBD residues shown to contact Su(H) (lane 2); wild-type S-CAP from the spider Parasteatoda tepidariorum (Ptep) (lane 3); Ptep S-CAP bearing the same five alanine substitutions (lane 4). Input levels of these His-tagged proteins for each experiment are shown in the respective “input” panels. Remaining two panels show the results of pulldown assays using Sepharose beads bearing only GST (left side) or GST-Su(H) (right side). Left: No binding of the His-tagged proteins to
GST alone is observed. Right: Wild-type Dmel Hairless binds efficiently to GST-Su(H) (lane 1); this interaction is severely reduced by the introduction of the five alanine substitutions (lane 2). Wild-type Ptep S-CAP likewise binds to GST-Su(H) (lane 3), while no binding is observed with the alanine-substitution mutant (lane 4); the same result is obtained even when the amount of input Ptep S-CAPs (wild-type and mutant) is increased by a factor of 10 (lanes 3 and 4, 10X). Amounts of GST and GST-Su(H) on the beads are shown in the Coomassie stains below the corresponding pulldown lanes. The results shown in this panel have been replicated in two additional experiments, including one utilizing new isolations of GST-Su(H) and His-tagged proteins.

Figure 5. Sequence characteristics of S-CAP proteins in Myriapods and Chelicerates. (A) Diagrams of representative examples of Myriapod and Chelicerate S-CAP proteins, denoting locations of SBD motifs and CtBP recruitment motifs. Scale and protein sizes are indicated. (B) Chelicerate, but not Myriapod, S-CAP proteins share N-terminal ELM2 and SANT domains with an MTA zinc-finger protein from the same species. Scale and protein sizes are indicated. (C) Shared ELM2 and SANT domains in Chelicerate MTA and S-CAP proteins are encoded in two exons with the same splice junction (indicated by /; red arrows). Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs. Species names as in Figure 4A.
Figure 5-figure supplement 1. Alignments of sequence regions shared by representative S-CAP proteins from (A) Chelicerates and (B) Myriapods. Amino acids at the ends of the aligned sequences are numbered. Pairwise amino acid sequence identities are indicated by vertical lines; conservative substitutions are indicated by + signs. ELM2 and SANT domains found in Chelicerate S-CAPs (A) are highlighted in colors corresponding to those in Figure 5B. Species names are as follows: (A) Hruf (Hypochthonius rufulus), Lpol (Limulus polyphemus), Otur (Ornithodoros turicata) (Egekwu et al. 2016), Tser (Tityus serrulatus) (Fuzita et al. 2015), Lhes (Latrodectus hesperus) (Clarke et al. 2015); (B) Smar (Strigamia maritima), Agir (Anopsobius giribeti).

Figure 6. Genes encoding both Hairless and S-CAP proteins are frequently located immediately adjacent to an MTA gene. Separate scale for each diagram is shown at the left. Three examples are shown for Hairless: the carpenter bee Ceratina calcarata (Ccal), the wheat aphid Schizaphis graminum (Sgra) (QEWZ01001380.1), and the lady beetle Harmonia axyridis (Haxy). Note that microsynteny is often preserved even when gene locations and relative orientations are changed. One example each is shown for S-CAP in Myriapods [the centipede Strigamia maritima (Smar)] and Chelicerates [the house spider Parasteatoda tepidariorum (Ptep)]. See also Supplementary file 1 and Supplementary file 4.

Figure 7. Speculative sources for key elements of the Hairless and S-CAP proteins. Note that these are intended to be only illustrative examples; other sources are of course
possible. (A) A C-terminal segment of the highly conserved Yippee-like protein (Roxström-Lindquist and Faye 2001; Hosono et al. 2004) is closely related to the C-terminal half of Hairless and S-CAP SBDs. Upper diagram is a sequence alignment of the entire Yippee-like proteins from *Drosophila melanogaster* (Dm) and *Homo sapiens* (Hs).

Aligned below are contiguous SBD motifs from *Drosophila* Hairless and five Myriapod and Chelicerate S-CAPs; their C-terminal halves are shown in bold. Two leucine (L) residues shown to make direct contact with Su(H) (Yuan et al. 2016) are highlighted in red. Amino acid sequence identities are indicated by vertical lines; conservative substitutions are indicated by + signs. Other species names as in Figure 4A. (B) As shown in the gene diagram at the bottom, the CtBP recruitment motif in Hairless is encoded by a very small exon located at the extreme 3' end of the gene [example is from the Oriental fruit fly *Bactrocera dorsalis* (Bdor; JFBF01000273.1); scale indicated]. A pre-existing gene encoding a protein that utilizes the same PLNLS recruitment motif is a possible source of this exon. Example shown is a portion of the *senseless* gene from the red flour beetle *Tribolium castaneum* (Tribolium GSC 2008). Senseless directly recruits the CtBP co-repressor via the PLNLS motif (Miller et al. 2014). This portion of the protein is encoded in exon 2; splice junction is indicated by a red /. Aligned beneath it is the last exon of the Bdor *Hairless* gene, illustrating its splice junction in the same frame as *senseless* exon 2.
Supplementary file legends

Supplementary file 1. Representative catalog of Hairless proteins in the Pancrustacea, selected from a curated collection of approximately 400 full-length sequences. Annotated sequences of the entries in this list are provided in Supplementary file 2. In the “MTA microsynteny?” column, + and – in parentheses denote the relative orientation of the Hairless and MTA genes in the genome.

Supplementary file 2. Annotated FASTA file of sequences of the Hairless proteins included in Supplementary file 1. Characteristic conserved domains and motifs are colored as in Figure 1A.

Supplementary file 3. Annotated FASTA file of sequences of the Su(H) proteins included in Table 1; species for which a full-length Su(H) sequence is not available are omitted here. Colors denote conserved sequence features. Motifs for direct recruitment of the CtBP and Gro co-repressor proteins (aligned in Table 1) are shown in red and green, respectively. Large region highlighted in orange is the highly conserved body of Su(H), extending from “LTREAM” to “YTPEP”.

Supplementary file 4. Representative catalog of S-CAP proteins in the Myriapoda and Chelicerata, selected from a curated collection of approximately 50 full-length sequences. Annotated sequences of the entries in this list are provided in
Supplementary file 5. In the “MTA microsynteny?” column, + and – in parentheses denote the relative orientation of the S-CAP and MTA genes in the genome. The “H SBD splice?” column indicates whether the Su(H)-binding domain (SBD) in the listed protein is encoded by two exons with the same splice junction as in Hairless. If not, the alternative exon structure is indicated.

Supplementary file 5. Annotated FASTA file of sequences of the S-CAP proteins included in Supplementary file 4. Characteristic domains and motifs are colored as follows: Su(H)-binding domain (SBD), orange; CtBP recruitment motifs, red.
Figure 1
Figure 3

A

<table>
<thead>
<tr>
<th>Smar Su(H) PVDL Smut</th>
<th>Smar Su(H) WT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>input protein</strong></td>
<td></td>
</tr>
<tr>
<td><strong>GST-dCtBP bound</strong></td>
<td></td>
</tr>
<tr>
<td>1 2 lane</td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>His-HLHmβ-WRP trunc</th>
<th>His-HLHmβ-WT</th>
<th>His-HLHmβ+Smar-motifMUT</th>
<th>His-HLHmβ+Smar-motifWT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pulldown: GST-Gro</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GST-Gro protein</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pulldown: GST</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GST protein</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 lane</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 4**
**Figure 5**

**A**

*Strigamia maritima* S-CAP (Myriapoda; Geophilomorpha)

*Anopsobius giribeti* S-CAP (Myriapoda; Lithobiomorpha)

*Parasteatoda tepidariorum* S-CAP (Arachnida; Araneae)

*Mesobuthus martensii* S-CAP (Arachnida; Scorpiones)

*Limulus polyphemus* S-CAP (Merostomata; Xiphosura)

**B**

*Parasteatoda tepidariorum* S-CAP (Arachnida; Araneae)

*Parasteatoda tepidariorum* MTA (Arachnida; Araneae)

*Strigamia maritima* MTA (Myriapoda; Geophilomorpha)

*Strigamia maritima* S-CAP (Myriapoda; Geophilomorpha)

**C**

Ptep MTA ELM2

Ptep S-CAP ELM2

Ptep MTA SANT

Ptep S-CAP SANT

Lpol MTAa ELM2

Lpol S-CAPa ELM2

Lpol MTAa SANT

Lpol S-CAPa SANT

---

**G/GC**

**G/GA**
**A**

| Dm CG15309  | MVK----TFQAYLPSTNRTYSVFHCRAHLSHDELISKSFSQGSQPGAYLFNSVNVACGQTEER |
| Hs YPEL1    | MVKMTKSSTTFOAYLPNHRTYSCIRCAHLSANHDELISKSFSQGSQGRAYLFNSVNVGCAPAEER |
| Dm CG15309  | VLLTGLHAVADRYCECKTLGWKHEAYESSQGYKEKFIIELAHMIKENGDWE |
| Hs YPEL1    | VLLTGLHAVADYCECNCKTLGWKHEAFESSQKYKEFIFIELAHMIKDNGWE |

Dm Hairless SBD  
Smar S-CAP SBD  
Agir S-CAP SBD  
Ptep S-CAP SBD  
Mmar S-CAP SBD  
Lpol S-CAPa SBD

---

**B**

*Tribolem castaneum senseless* (exon 2)

/ATGAGGAGGCGLCCCCCTCAACCTTCCTCCCTCAAGGGCCGTCGCGCAGCCAC...

E E A P L N L S L K G R S R S R S H ...  

/ACGTGCCGCTGAACTTTCTCTGACTAGCATTTGA  

V P L N L S K H *  

*Bactrocera dorsalis Hairless* (last exon)

---

![Figure 7](Figure 7)
Figure 1-Figure Supplement 1
Alignment of Chelicerate cS-CAP Conserved Domains

Hruf S-CAP ELM2 21 PNRNRIR1GRQYQATVPPLLKGESDGRKLEELTRWKPNDLSDQQLDQYLSVAK-AVSLFA 83
Lpol S-CAPa ELM2 20 KNRNRIR1VGRQYQATVPPLLKGESDGRQLEDLETLEKWPNDLSDQQLDQYLSVAR-AVSFVA 82
Otur S-CAP ELM2 19 NNRNRIR1VGRQYQATVPPLLKGESDGRRLDELETLEKWRPES-LSQDITIVEYMLSAM-AVSLFA 80
Tser S-CAP ELM2 25 RNRNRIR1VGRQYQATVPPLLKGEP1DGCEDLETLEKWPNDLSDQIDQYLSMASRAVSLFA 88
Lhes S-CAP ELM2 21 QNRNRIR1VGRQYQATIPPLLKGKEKDGRKCEDLETLEKWEQ-LSQDKEEYMSAM-GTSLF 82

Hruf S-CAP 133 SASVGDHIQSALGKLSDFVTSHHPGHHDAGCRVSDP 169
Lpol S-CAPa 91 SDKESDKC1QTALRGLTEVTSHHPCHHDANCQLSOG 127
Otur S-CAP 89 SEEDSGC1QTALKLSDFVTSHHGCHHDACCGWGP 125
Tser S-CAP 97 TENGPD1KTLQRGLSDFVTSHHPCHHDACCGVPS 133
Lhes S-CAP 90 TSEKDSDLQSIAKGLTEFVASHHYHDDGCRVFK 126

Hruf S-CAP SANT 185 SGLITGEWTPNEMKLTPRALVDCGKNFGAIKKEFLPWKSVSIIHYYLGIDKLLS 240
Lpol S-CAPa SANT 135 GSNNSSNDWTYGAQLPARALEACGKNFGAIKKEFLPWKSVSIIHYYGKEEPRN 190
Otur S-CAP SANT 123 GPLEPSDWTPEANSLARALEACGKNSFAIKKDELPWKLKVSLIYYYGKLEKQD 178
Tser S-CAP SANT 142 KHSIKNGWTPTEAHLFARALEACGKNFGAIKKEFLPWKPVRSIIHYYNGKEEKYD 197
Lhes S-CAP SANT 135 PKISTNWSPFEAQLFQALESCGKNFCAIKKEFLPWKPVRSIIHYYQIKNKNE 190

Alignment of Myriapod mS-CAP Conserved Regions

Smar S-CAP 262 PDNASLPLPEWYLDPSKSQIPVNEIGNPRPYPALAPTH 300
Agir S-CAP 534 ASSSNLPLPEWYLDPSKAEPVSEMIYGPRPYPALPPSTD 572

Smar S-CAP 361 PSADVQFWPCVSCPVCVAYAPTETTVNSYYYCCPLPGCRSNCLKTEGKQLTS 410
Agir S-CAP 1009 SIGLAPFPYWPCPPVCVAYGPAVNSYYYCCPLPGCRHMYMPREPVI 1058

Smar S-CAP 454 DGCSSMQLKDLMLMPSS 471
Agir S-CAP 1139 SATSGSMKLKDLMPSSRT 1156

Figure 5 - Figure Supplement 1