Testis-specific TAF homologs collaborate to control a tissue-specific transcription program

Mark Hiller*, Xin Chen, M. Jodeane Pringle, Martin Suchorolski, Yasemin Sancak, Sridhar Viswanathan, Benjamin Bolival, Ting-Yi Lin, Susan Marino and Margaret T. Fuller†

Departments of Developmental Biology and Genetics, Stanford University School of Medicine, Stanford, CA 94305-5329, USA
*Present address: Department of Biological Science, Goucher College, Baltimore, MD 21204, USA
†Author for correspondence (e-mail: fuller@cmgm.stanford.edu)

Accepted 17 June 2004

Development 131, 5297-5308
Published by The Company of Biologists 2004
doi:10.1242/dev.01314

Summary

Alternate forms of the PolII transcription initiation machinery have been proposed to play a role in selective activation of cell-type-specific gene expression programs during cellular differentiation. The *cannonball* (*can*) gene of *Drosophila* encodes a homolog of a TBP-associated factor (dTAF5) protein expressed only in spermatocytes, where it is required for normal transcription of genes required for spermatid differentiation. We show that *Drosophila* primary spermatocytes also express four additional tissue-specific TAFs: *nht* (homolog of dTAF4), *mia* (homolog of dTAF6), *sa* (homolog of dTAF8) and *rye* (homolog of dTAF12). Mutations in *nht, mia* and *sa* have similar effects in primary spermatocytes on transcription of several target genes involved in spermatid differentiation, and cause the same phenotypes as mutations in *can*, blocking both meiotic cell cycle progression and spermatid differentiation. The *nht, mia, sa* and *rye* proteins contain histone fold domain dimerization motifs. The *nht* and *rye* proteins interact structurally when co-expressed in bacteria, similarly to their generally expressed homologs TAF4 and TAF12, which heterodimerize. Strikingly, the structural interaction is tissue specific: *nht* did not interact with dTAF12 and dTAF4 did not interact with *rye* in a bacterial co-expression assay. We propose that the products of the five *Drosophila* genes encoding testis TAF homologs collaborate in an alternative TAF-containing protein complex to regulate a testis-specific gene expression program in primary spermatocytes required for terminal differentiation of male germ cells.

Key words: TAF, TFIID, *Drosophila*, Spermatogenesis, Spermatocyte, Transcription

Introduction

The initiation of transcription by RNA polymerase II (PolII) remains a critical control point for regulation of differential gene expression during development and the differentiation of specialized cell types. The general transcription factor TFIID is thought to play a central role in interpreting and integrating molecular signals regulating the core PolII transcription machinery for initiation (Hochheimer and Tjian, 2003). Recruitment of TFIID to sequences near the transcriptional start site is thought to in turn recruit and/or stabilize PolII binding in the preinitiation complex (Lemon and Tjian, 2000; Orphanides et al., 1996; Roeder, 1996). The multisubunit TFIID complex contains the TATA-binding protein (TBP) and 12-14 TBP-associated factors (TAFs) (Albright and Tjian, 2000). Recently, alternative forms of TBP and tissue-specific TAFs have been described in a number of organisms (reviewed by Hochheimer and Tjian, 2003; Veenstra and Wolffe, 2001), raising the possibility that cell-type- or stage-specific forms of what was previously thought of as the general transcription machinery may play an important role in selective activation of certain PolII promoters (Verrijzer, 2001). To date, however, only a few tissue-specific TAF_{II}s have been investigated for in vivo function in cell-type-specific gene expression programs. For example, an alternate form of TAF4 (TAF4b – formerly $hTAF_{II}105$) highly expressed in granulosa cells was found to associate with TBP and TAF1 in a large, TFIID-like protein complex in ovarian extracts and to be required for normal follicular development and for expression of a number of genes in developing ovarian follicles in mice (Freiman et al., 2001).

A striking example of a tissue-specific TAF homolog required for execution of a developmentally regulated transcriptional program appears during differentiation of male germ cells in *Drosophila* (Hiller et al., 2001). The *cannonball* gene (*can*) of *Drosophila*, which encodes a homolog of dTAF5 expressed only in male germ cells, is required for normal transcription in primary spermatocytes of a set of spermatid differentiation genes required for normal spermatogenesis (Hiller et al., 2001; Lin et al., 1996; White-Cooper et al., 1998). Flies null mutant for *can* are viable and female fertile but male sterile. The requirement for *can* function is gene selective: only a specific set of genes normally expressed in wild-type primary spermatocytes are affected, while a number of other genes are transcribed normally in spermatocytes from *can* mutant males. While it is now clear that cell-type-specific TAF homologs such as *can* and mTAF4b can play important roles in tissue-specific gene expression, the mechanisms by which they function at specific promoters are not understood. To identify

proteins that might collaborate with the dTAF5 homolog *can* to regulate expression of specific target genes in *Drosophila* spermatocytes, we investigated the expression and function of other TAF_{II} homologs in the *Drosophila* genome.

We show that *Drosophila* primary spermatocytes express several additional tissue-specific TAF homologs that act to control expression of spermatid differentiation genes: *nht* (homolog of dTAF4), *mia* (homolog of dTAF6), *sa* (homolog of dTAF8), and *rye* (homolog of dTAF12). Mutations in *nht*, *sa* and *mia* cause the same phenotypes as *can*, blocking meiotic cell cycle progression and spermatid differentiation, and have similar effects on transcription in primary spermatocytes of several target genes involved in spermatid differentiation. The *nht* and *rye* proteins interact structurally when co-expressed in bacteria, like their generally expressed homologs TAF4 and TAF12. Strikingly, the structural interaction was tissue-specific: *nht* did not interact with dTAF12, and dTAF4 did not interact with *rye* in the bacterial co-expression assay. We propose that these five *Drosophila* testis-specific TAF homologs collaborate in an alternative TAF-containing protein complex expressed in primary spermatocytes to regulate expression of a set of genes involved in spermatid differentiation.

Materials and methods

Drosophila strains, and husbandry, deficiency complementation

Drosophila were raised on standard cornmeal and molasses medium at 25°C. mia^{1} has been described by Lin et al. (Lin et al., 1996). $mia^{z\text{-}3348}$, $nht^{z\text{-}5347}$ and $nht^{z\text{-}5946}$ were isolated in the EMS-induced male sterile screen (Wakimoto et al., 2004). Agametic flies were generated using osk^{8} and osk^{ce3} lines as described by Hiller et al. (Hiller et al., 2001). Deficiency mapping, viability tests and fertility tests were conducted as described previously (Lin et al., 1996). Deficiencies are described in FlyBase or by Lin et al. (Lin et al., 1996). *nht* mutants failed to complement *Df(2L)osp38, Df(2L)TE35BC-4* and *Df(2L)A263*, but complemented *Df(2L)A72, Df(2L)TE35D-21, Df(2L)TE35D-2* or *Df(2L)noc11* for male sterility and the meiotic arrest phenotype. *sa* alleles failed to complement *Df(3L)Pc-MK, Df(3L)Pc-CP1, Df(3L)Pc-12h, Df(3L)Pc-14d* and *Df(3L)Pc-2q*, but complemented *Df(3L)Pc* or *Df(3L)Pc-101* for male sterility and the meiotic arrest phenotype. Mapping breakpoints of *Df(3L)Pc-MK* and *Df(3L)Pc* at the molecular level by single embryo PCR narrowed down *sa* to a 72 kb genomic region. Mobilization of P-element insertion line *EP(3)0339* (Szeged Stock Center) as described by Tower et al. (Tower et al., 1993) generated lesions within this 72 kb by imprecise excision. *sa* was mapped proximal to the *EP(3)0339* insertion based on 12 such lines, which were initially detected by lethality with existing deficiencies in the region, then tested for complementation with *Df(3L)Pc-MK, Df(3L)Pc* and two *sa* alleles.

Molecular cloning and sequence analysis of *nht*, *mia*, *sa* and *rye*

cDNAs corresponding to *mia*, *nht* and *rye* transcripts were isolated by screening a testis cDNA λZAP library provided by T. Hazelrigg (Columbia University) or the BDGP AT library (http://www.fruitfly.org/EST/index.html). cDNA corresponding to the *sa* coding region was obtained by RT-PCR. A 79-base 5′ UTR for *sa* and a possible 5′ end of the mRNA for *nht* and *rye* were identified from testis poly(A)$^{+}$ mRNA using the 5′ GeneRacer system (Invitrogen). No different forms of the *mia* predicted protein coding region were apparent from the ESTs in the BDGP database, the two testis cDNA clones characterized, or RT-PCR of fragments of the *mia* ORF from testis mRNA compared to mRNA from 0-24-hour-old embryos. Primers for RT-PCR corresponded to base pairs 956-977 and bp 1805-1825 of the cDNA, residing in exon 1 and 3, respectively. RT-PCR using embryo RNA gave an 870 bp product, the expected size for the fully processed *mia* transcript target, as well as a 1320 bp product expected if unprocessed *mia* transcript was the template. No products were generated in control reactions lacking RT.

The *mia* genomic rescue construct contained a 6.1 kb *Eag*I-*Xho*I genomic fragment from cosmid 80E1 (Clare O'Connor, Boston College) cloned into the *Not*I-*Xho*I sites of pCaSpeR-4 (Rubin and Spradling, 1983). The approximately 10.5 kb *nht* genomic rescue construct D7 is described by Yagi and Hayashi (Yagi and Hayashi, 1997). A 3.2 kb genomic DNA containing 1 kb both upstream and downstream of the *sa* coding sequence obtained by PCR from wild type genomic DNA was cloned into pCasper-4 and introduced into flies by P-element transformation. The *mia*, *nht* or *sa* genomic transgenes rescued the respective mutant males to fertility and restored normal spermatogenesis, based on phase contrast microscopy of testis squash preparations.

RT-PCR to assess tissue-specific expression of *sa* was performed using Qiagen Onestep RT-PCR kit with primers corresponding to the first 18 and the last 18 nt of the predicted protein coding sequence producing an 825 bp fragment amplified from male RNA. Point mutations in the *nht*, *mia* and *sa* alleles were identified by sequencing bulk PCR products amplified from genomic DNA of homozygous mutant flies using gene-specific oligonucleotides. cDNA and PCR sequences were aligned and analyzed using Sequencher (Gene Codes Corp.) and MacVector (Oxford Molecular Group plc) DNA analysis software.

The Nht, Mia, Sa and Rye protein sequences, predicted from the cDNA sequences, were used to search nucleotide sequence databases translated in all reading frames (tBLASTn). The cDNA sequences were AY752877(nht), AY752878(mia), AY752879(sa) and AY752880(rye).

Alignments of *nht*, *rye*, *mia* and *sa* predicted proteins to the histone fold domains of the generally expressed TAFs were constructed using HMMER release 2.3 (Durbin et al., 1998) against the PFAM database V.10.0 (Bateman et al., 2002). These domains were used to seed multiple sequence alignments performed using ClustalW (Thompson et al., 1994) and the T-Coffee multiple sequences alignment method (Notredame et al., 2000). The multiple alignments were performed through stepwise, independent TAF and histone family alignments then by aligning separate profiles of each family. With the highly divergent sequences, T-Coffee was found to provide much more reliable alignments than ClustalW, and was used in the alignments presented in this paper. Alignment coloring and percentage identity/similarity for pairwise comparisons were obtained using GeneDoc (Nicholas et al., 1997).

Protein alignments were refined and threaded onto existing x-ray crystal structures using the Swiss-PDB Viewer software package and Swiss-Model server (http://www.expasy.org/swissmod/SWISS-MODEL.html) (Guex and Peitsch, 1997). Putative structures of *nht* and *rye* were threaded onto 1H3O. Protein alignments were modified using the Sippl-like threading energy minimization functions in Swiss-PDB Viewer and resubmitted to the Swiss-Model server until optimal RMSD was obtained. An average noise cut-off RMSD of 0.87A was obtained by threading and averaging the RMSD of three gap-free helix-loop-helix proteins unrelated to the TAF family. As negative controls, we also aligned, as oligomers, Nht-TAF12 and TAF4-Rye.

In situ hybridization and RNA and DNA blot analysis

In situ hybridization and DNA/RNA blot analysis were as described previously (Hiller et al., 2001). mRNA from adult testis or 0- to 24-hour embryos was isolated by homogenization in TRIzol Reagent (Life Technologies) followed by isolation of mRNA using Micro-FastTrack 2.0 (Invitrogen). RT-PCR of *mia* was carried out using

Qiagen OneStep RT-PCR kit. For northern blots and in situ hybridization, the *mia* probe was generated from a DNA fragment corresponding to the first 1331 bp of the *mia* cDNA, the *nht probe* was generated from a plasmid containing a PCR product corresponding to bp 89-789 of the cDNA, and the *rye* probe was generated from a plasmid containing PCR sequences corresponding to genomic DNA, including intron sequences, between bp 16 and 438 of the cDNA. For *sa*, in situ hybridization was performed with antisense and sense probes of the entire protein coding sequence. Other probes for in situ hybridization have been described previously (White-Cooper et al., 1998; Hiller et al., 2001).

Tests of physical interaction in bacteria

The GST-Nht fusion protein was produced by introducing *Eco*RI (5′ region coresponding to amino acid 5) and *Hin*dIII (immediately 3′ to the natural stop codon) sites into sequences flanking the Nht coding region and cloning that region into vector pGEX-KG (Guan and Dixon, 1991). A similar region of dTAF4, amino acid 647 through the natural stop codon, was also cloned into pGEX-KG. A portion of the Rye protein was expressed from pACYC184-11b by cloning a PCR product encoding amino acid 65 to the natural stop codon. Primers used to generate the PCR product were designed to introduce a Met residue preceeding the first Ile (amino acid 65) of the Rye sequence, followed by Pro Tyr Phe Ser Pro Tyr Gln. A similar dTAF12 fusion protein expressed amino acids 91-169. A linker containing the FLAG epitope (Met Tyr Lys Asp Asp Asp Asp Lys Ala Ala Ala) was inserted before the first Met (converted *Nde*I site) of the Rye and dTAF12 fusion proteins. All constructs were sequenced to insure that no unintended changes were introduced by PCR and fusion joints were in-frame.

Co-expression of histone-fold-containing fusion proteins made use of expression vectors pGEX-KG (Guan and Dixon, 1991) to generate appropriate GST fusion proteins, and pACYC184-11b (Fribourg et al., 2001), which contains a p15A origin of replication compatible with the ColE1 origin of replication on pGEX-KG, to generate FLAG-tagged proteins. Plasmids were transformed into bacterial strain BL21, and grown at 26°C in medium containing 100 μg/ml ampicillin and 35 μg/ml chloramphenicol to an A600 of 0.5. Expression of fusion proteins was induced with 0.3 mM isopropyl β-D-thiogalactoside (IPTG) and followed by growth overnight at 16°C. GST-fusion proteins were purified using glutathione-Sepharose 4B (Pharmacia Biotech, Inc) per the manufacturer's instructions. Proteins were separated by SDS-PAGE and transfered to nitrocellulose membrane (Hybond ECL, Amersham Pharmacia Biotech). Blots were probed with anti-GST (Amersham Pharmacia) at a dilution of 1:1000 and anti-FLAG M1 (Sigma) at 1:666. In our studies, as also shown by several other groups working with HFD TAFs, proteins with a histone-fold motif were often not soluble in the absence of their proper binding partner. As purification by binding to glutathione-Sepharose requires the GST fusion proteins to be stable and soluble, insoluble proteins were not detected in our binding assays. The GST-TAF4 fusion protein fragment we tested was slightly soluble when expressed alone and could be detected upon longer exposures. The TAF12-FLAG fusion protein we tested was stable, but did not bind and elute from the glutathione beads in the absence of its TAF4 HFD-containing binding partner. Solubility and co-purification of FLAG-Rye on glutathione-Sepharose was only observed in the presence of GST-Nht. Likewise, purification of FLAG-TAF12 on glutathione-Sepharose was only observed in the presence of GST-TAF4.

Results

The *Drosophila* genome encodes testis-specific homologs of several generally expressed TAFs

Five genes encoding predicted second homologs of biochemically identified *Drosophila* embryo TAFs were found encoded in the *Drosophila* genome by tBlastn searches (Table 1) (Adams et al., 2000). In addition to *cannonball* (*can*), which encodes a homolog of dTAF5 (Hiller et al., 2001) the *Drosophila* genome contains predicted second homologs of several other TAFs: *no hitter* (*nht*) (CG15259) a homolog of TAF4, *spermatocyte arrest* (*sa*) a homolog of dTAF8 (CG11308), *meiosis I arrest* (*mia*) (CG10390) (see also Aoyagi and Wassarman, 2000) a homolog of TAF6 and *ryan express* (*rye*) (CG15632) a homolog of TAF12 (Table 1). In addition, the *Drosophila* genome encodes two previously described TAF10 homologs, dTAF10 and dTAF10b, both expressed during embryogenensis (Hernandez-Hernandez, 2001; Georgieva, 2000).

The newly identified TAF homologs *nht, sa, mia* and *rye* are expressed either exclusively or mainly in the testis. The 1 kb *nht* and the 0.6 kb *rye* transcript were detected by northern blot analysis in adult males but not in adult females, adult males lacking a germline, or embryos (Fig. 1A). Likewise, RT-PCR indicated expression of *sa* mRNA in adult males and in testes, but not in adult males lacking germline, females or embryos (Fig. 1B). For *mia*, northern blots revealed a predominant 3 kb transcript in adult males but not in agametic males, females or embryos (Fig. 1A). In addition, one or two smaller transcripts expressed at much lower levels were detected in females and embryos in northern blots probed for *mia*. Expression of *mia* in embryos was confirmed by RT-PCR and *mia* ESTs were identified from embryo- as well as testis-derived libraries in the BDGP EST sequence database.

Within the testis, transcripts for *nht*, *sa*, *rye* and *mia* were detected in primary spermatocytes upon in situ hybridization of an antisense RNA probe to whole-mount testes (Fig. 1C-F). *nht*, *sa*, *rye* and *mia* transcripts appeared first in male germ cells at the transition from the spermatogonial stage to the primary spermatocyte growth phase (arrows). Like *can* mRNA, the level of *nht*, *sa*, *rye* and *mia* mRNAs detected was characteristically highest in early spermatocytes, with the signal decreasing in more mature spermatocytes. *nht*, *sa*, *rye* and *mia* transcripts were not detected by in situ hybridization

Table 1. Gene names

Generally-expressed TAF* (polytene)	Testis-expressed homolog (polytene)	Testis-expressed homolog gene name [CG number]
dTAF4 (72D)	dTAF4L (35C)	*no hitter* (*nht*) [CG15259]
dTAF5 (47C)	dTAF5L (65D)	*cannonball* (*can*) [CG6577]
dTAF6 (76B)	dTAF6L (83B)	*meiosis I arrest* (*mia*) [CG10390]
dTAF8 (*Prodos*) (16F7)	dTAF8L (78D1)	*spermatocyte arrest* (*sa*) [CG11308]
dTAF12 (86E)	dTAF12L (25A)	*ryan express* (*rye*) [CG15632]

*According to the recently revised nomenclature for TAFIIs (Tora, 2002), the *Drosophila* TAF homologs are designated dTAF4L, dTAF5L, dTAF6L, dTAF8L and dTAF12L, as they have not yet been demonstrated to participate in a complex with TBP.

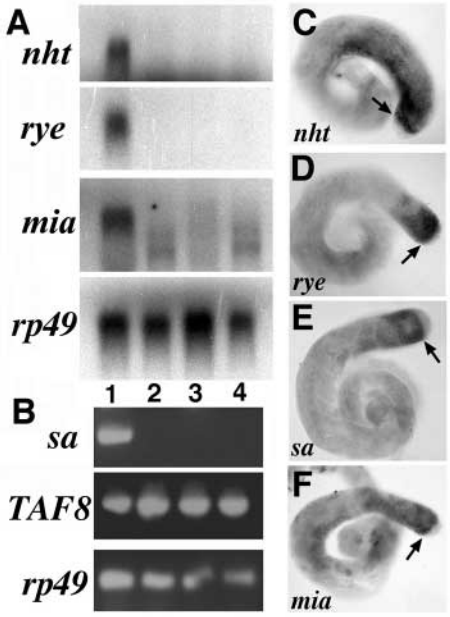


Fig. 1. Expression of the testis TAF homologs *nht, rye, mia* and *sa*. (A) Northern blot of poly(A)$^+$ selected RNA probed sequentially with sequences from the protein coding regions of *nht, rye* and *mia*, and with *rp49* as loading control. (B) RT-PCR products amplified using primer pairs for *sa*, its generally expressed *Drosophila* homolog *dTAF8*, and *rp49* as loading control from Poly(A)$^+$ RNA. In both A and B, RNA was from: wild-type adult males (lane 1), wild-type adult females (lane 2), adult males lacking germline (lane 3), 0- to 24-hour embryos (lane 4). (C-F) *nht, rye, sa* and *mia* mRNAs were expressed in primary spermatocytes. In situ hybridization to wild-type whole adult *Drosophila* testis with single stranded antisense riboprobes: (C) *nht*, (D) *rye*, (E) *sa*, or (F) *mia*. Arrows indicate the onset of mRNA expression at the start of the primary spermatocyte stage.

in the apical tip region of the testes containing stem cells and spermatogonia.

The testis TAF homologs mediate a common gene expression program in spermatocytes

To assess the role of the TAF homologs in vivo, we identified null mutant alleles. Two *nht* mutants were identified by screening a large collection of male sterile lines (Wakimoto et al., 2004) for mutants with a meiotic arrest phenotype similar to *can*. The mutation causing the *nht* meiotic arrest phenotype was localized to polytene region 35C by recombination mapping and deficiency complementation (Materials and methods). As this region contained CG15259, the dTAF4 homolog, we sequenced genomic DNA amplified from CG15259 by PCR. Both *nht* alleles carried base changes from the background chromosome that caused premature stop codons in the predicted CG15259 open reading frame (Table 2). A transgene with a 10.5 kb fragment of genomic DNA containing CG15259 fully rescued the spermatocyte arrest phenotype and male sterility of $nht^{z\text{-}5347}/Df$. Together these results established that *nht* encodes the TAF4 homolog (Table 1).

A second allele of *mia* was identified in the same phenotypic screen. *mia* had previously been mapped to either polytene interval 78C9-81F or 83A-C2 (Lin et al., 1996). As 83A-C2 contained the dTAF6 homolog CG10390, we sequenced genomic DNA corresponding CG10390 from the two *mia* alleles. mia^1 and $mia^{z\text{-}3348}$ had base changes resulting in stop condons, at aa297 and aa55, respectively (Table 2). A 6.1 kb fragment of genomic DNA containing CG10390 rescued the male sterility and spermatogenesis phenotype of *mia* mutants.

The *sa* locus was previously mapped to polytene chromosome interval 78C9-D2 (Lin et al., 1996). Additional local deletions made by mobilization of the *EP(3)0339* P-element insert (Materials and methods) further localized *sa* proximal to *EP(3)0339*. The *sa* gene was found to lie in a large intron of CG6014, transcribed from the opposite strand. Sequencing of genomic DNA amplified from the two *sa* alleles by PCR reveled a 414 bp deletion and 158 bp insertion in the predicted protein coding region in sa^1, which had been recovered from the wild as *VO45* by D. Lindsley in the Rome screen (Sandler et al., 1968), and a C to T transition resulting in a premature stop codon in the EMS-induced sa^2 compared to its background chromosome (Table 2). A 3.2 kb genomic DNA fragment containing CG11308 fully rescued the spermatocyte arrest and male sterility phenotypes of sa^1/sa^2 males when introduced into flies by P-element-mediated germline transformation, establishing that *sa* corresponds to CG11308 (Table 1).

Mutations in *nht, mia* or *sa* caused the same meiotic arrest phenotype as loss of function of the testis-specific dTAF5 homolog *can*. Testes from *nht* males contained cells at the earliest stages of spermatogenesis up through primary spermatocytes, but germ cells failed to initiate meiotic cell division and the testes filled with mature primary spermatocytes (Fig. 2B). No stages of spermatid differentiation were detected in *nht* testes and mature spermatocytes eventually degenerated at the testis base. Loss of function of *mia* or *sa* caused a similar meiotic arrest phenotype, with

Table 2. Mutations in *nht, mia* and *sa* alleles

Mutant allele	Base pair change*	Amino acid change†
$nht^{z\text{-}5347}$	C(566) to A	S(166) to stop
$nht^{z\text{-}5946}$	C(114) to T	Q(16) to stop
mia^1	G(1448) to A	W(297) to stop
$mia^{z\text{-}3348}$	G(772) to A	W(55) to stop
sa^1	414bp (145-558) deletion plus a 158bp insertion	Frameshift after the 47th aa
sa^2	C(326) to T	Q(109) to stop

*The position of the base pair change is given relative to the start of transcript as determined by a combination of the sequence of the cDNA and 5′ RACE products.

The position of amino acid changes are given relative to the predicted start of translation.

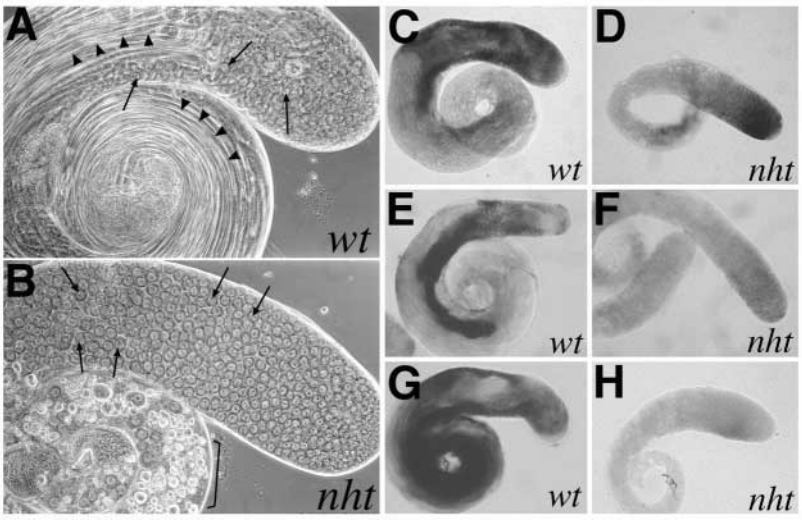


Fig. 2. Loss of function of *nht* arrests mature primary spermatocyte differentiation and lowers expression of spermatid differentiation genes. (A,B) Phase-contrast images of whole testes dissected from (A) wild-type and (B) *nht* mutant males. Arrows indicate primary spermatocytes; arrowheads, elongated spermatid bundles in wild-type; the bracket in B shows the region of dying spermatocytes in *nht*. (C-H**)** In situ hybridization to whole (C,E,G) wild type or (D,F,H) $nht^{z\text{-}5347}/Df(2L)A263$ mutant testes with anti-sense mRNA probes for (C,D) *cyclinB*, (E,F) *fuzzy onions*, (G,H) *don juan*.

mature spermatocytes arrested at the G2/M transition of meiosis I and lack of spermatid differentiation (Lin et al., 1996). Strikingly, although *mia* mRNA expression was not specific to the adult male germline, *mia* function appeared to be required mainly for spermatogenesis: $mia^{1}/mia^{z\text{-}3348}$ flies were viable and female fertile.

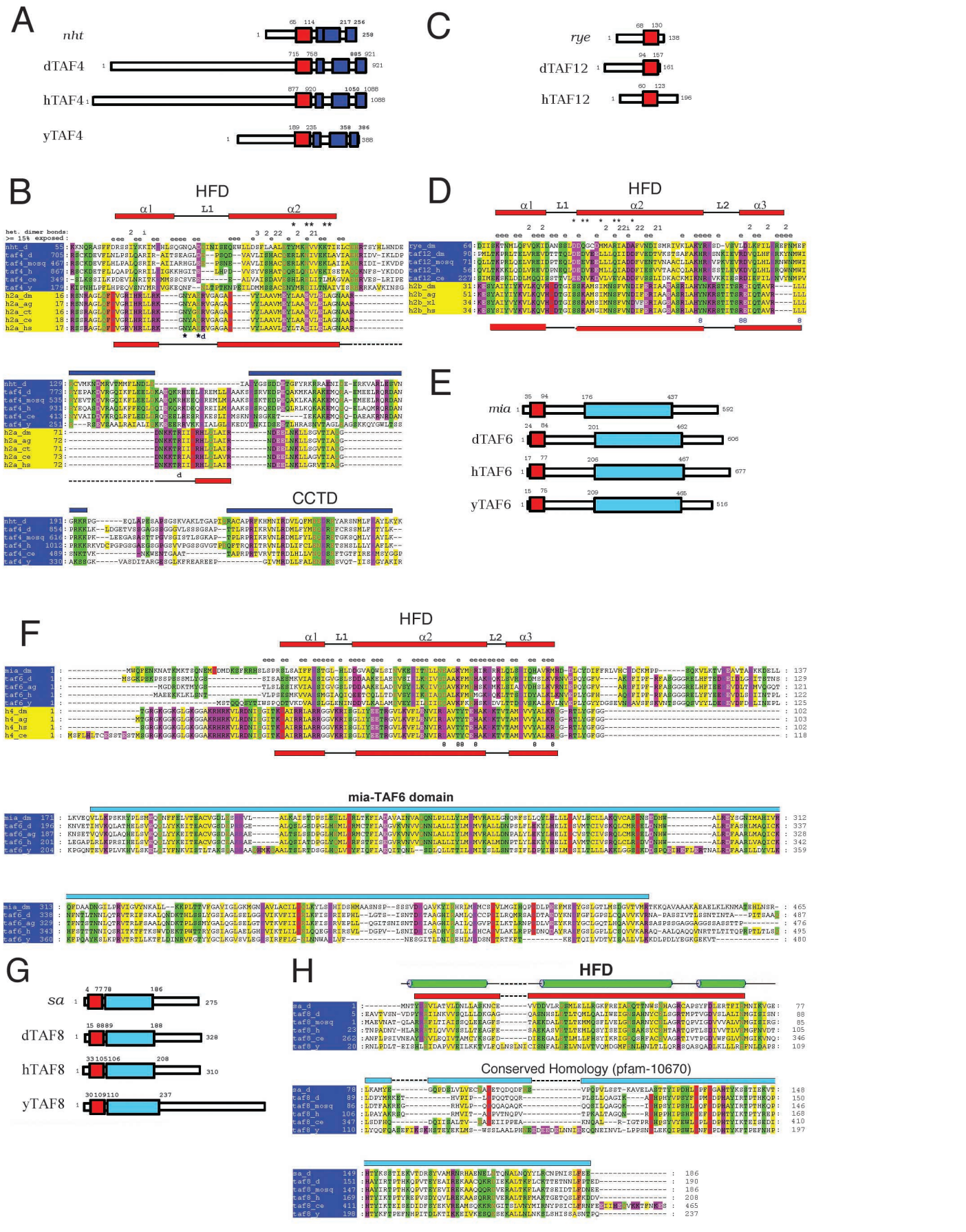
Like *can*, loss of function of *nht*, *mia* or *sa* caused defects in expression of spermatid differentiation genes in primary spermatocytes. Transcripts from the *can* target genes *fuzzy onions* (*fzo*) and *don juan* (*dj*) were expressed at much lower levels in spermatocytes in *nht* mutant testes compared to wild type, based on in situ hybridization using RNA probes (Fig. 2E-H). Similar results were obtained for other target genes that also require *can* for normal levels of expression in primary spermatocytes, including *janB* and *mst35B* (data not shown). As observed for *can*, although the level of target gene transcripts were significantly reduced in *nht* mutant testes, low levels of target gene messages were still detected by northern blot analysis of poly(A)$^{+}$ testis RNA (data not shown). Also as observed for *can*, wild-type function of *nht*, *mia* or *sa* was not required for expression of all spermatocyte transcripts: mRNAs for the meiotic cell cycle regulators *cyclin B* and *twine* (a *cdc25* homolog) were expressed normally in *nht* mutant testes, based on both in situ hybridization and northern blots (Fig. 2D and data not shown). Loss of function of *mia* or *sa* had previously been shown to have similar effects on the expression of the same genes as *can* (White-Cooper et al., 2000). With respect both to levels of target transcripts and effects on expression of *twine* and *cyclin B* transcripts in primary spermatocytes, *nht* was clearly like *can*, *mia* and *sa* rather than like the *aly* class meiotic arrest genes *aly, comr, topi* and *achi/vis* (Ayyar et al., 2003; Jiang and White-Cooper, 2003; Perezgasga et al., 2004; Wang and Mann, 2003; White-Cooper et al., 2000).

Several testis TAFs are homologs of general TAFs containing a histone fold domain motif

The full-length protein encoded by *nht* was homologous to the C-terminal portion of dTAF4 and its human homolog hTAF4, but lacked the characteristic N terminal glutamine rich domain (Fig. 3A). A 924 bp cDNA representing the *nht* transcript isolated from a testis cDNA library and confirmed by RT-PCR from testis mRNA closely matched the 1 kb size predicted by northern analysis. Sequencing of this near full length *nht* cDNA revealed an open reading frame encoding a predicted 245 amino acid protein. Comparison with the genomic DNA sequence revealed no introns in the region covered by the cDNA. The initial methionine was preceded by in frame stop codons in the cDNA, suggesting that it represented the start of the protein. The yeast homolog Mpt1p (yTAF4) also does not contain the extended N-terminal glutamine rich domain.

The C-terminal domain of TAF4 contains a histone fold domain (HFD), a dimerization domain present in several different TAF subunits. Through this characteristic α-helix – loop – α-helix – loop – α-helix structure (α1-L1-α2-L2-α3) several TAFs form specific heterodimer pairs, much like the histones H2A:H2B and H3:H4 (Werten et al., 2002; Xie et al., 1996). Although amino acid sequence conservation was relatively low (Table 3), the predicted Nht protein contained a domain (aa64-aa114 of *nht*) that could be aligned with the α1-L1-α2 region of the histone fold motif of TAF4 and histone H2A (Fig. 3B, top set). The predicted alignment for the *Drosophila* TAF4 homologs was checked by threading the amino acid sequences on to the crystal structure obtained for the hTAF4:hTAF12 heterodimer (Werten et al., 2002). The dTAF4 sequence fits very well, with an RMSD of 0.14A for the 715-758aa region. The *nht* sequence also fit into the structure, although the *nht* sequence was more diverged than dTAF4 from hTAF4 and the RMSD was considerably higher (0.74A, even after optimizing by comparing RMSDs for three different possible local alignments).

The homology between the predicted *nht* protein and TAF4 proteins from yeast to humans was particularly striking in the well conserved pattern of alternating hydrophobic and charged or polar residues in the region of alpha helix 2, which forms the core of the interaction domain between TAF4 and its heterodimer partner TAF12. The region of hTAF4 that might contribute to the predicted alpha 3 helix was not included in the crystal structure and is the subject of some debate (Thuault et al., 2002; Werten et al., 2002). However, the predicted *nht* protein did share with the TAF4 homologs from human to yeast a characteristic conserved pattern of hydrophobic, charged and polar residues in the CCTD region near the C terminus of the


A
nht
dTAF4
hTAF4
yTAF4
C
rye
dTAF12
hTAF12
B
HFD
α1
L1
α2
CCTD
D
HFD
α1
L1
α2
L2
α3
E
mia
dTAF6
hTAF6
yTAF6
F
HFD
α1
L1
α2
L2
α3
mia-TAF6 domain
G
sa
dTAF8
hTAF8
yTAF8
H
HFD
Conserved Homology (pfam-10670)

proteins (Fig. 3B, bottom set). In addition, multiple sequence alignments by T-Coffee (Notredame et al., 2000) revealed two other extended regions between the HFD and the CCTD that had patterns of charged, polar and hydrophobic residues conserved from *nht*, to its *Drosophila*, mosquito, human and *C. elegans* homologs (Fig. 3B, middle set).

The predicted protein encoded by *rye* is homologous to TAF12, the histone fold motif-containing binding partner of TAF4. A 550 bp *rye* cDNA isolated from a testis cDNA library closely matched the 600 bp transcript size estimated from northern blots (Fig. 1). Sequencing of this near full length *rye* cDNA revealed an open reading frame encoding a 138aa predicted protein with homology to dTAF12. An in frame stop codon located six base pairs upstream of the predicted initial methionine in the cDNA suggested that this corresponded to the start of the protein.

The predicted Rye protein contained a histone fold domain region (amino acid residues 68-130 of *rye*) that aligned well with the histone fold domain of TAF12 and histone H2B (Fig. 3C,D). The sequence of the generally expressed dTAF12 as aligned in Fig. 3D threaded onto the crystal structure of hTAF12 in the hTAF4:hTAF12 dimer well, with an RMSD of 0.09A for the region from aa91-aa164. Although *rye* was more diverged, the *rye* HFD sequence aligned as in Fig. 3D also fit onto the hTAF4:hTAF12 dimer crystal structure (RMSD=0.47). Again, *rye* and the TAF12 proteins from *C. elegans* to humans showed a striking pattern of alternating hydrophobic and charged or polar residues echoing the similar pattern in histone H2B in the region of the predicted HFD, especially in the second helix (Fig. 3D).

The *mia* gene encodes a protein homologous to dTAF6. A 2400 bp cDNA isolated from a testis cDNA library corresponded in size to the *mia* transcript expressed in testis estimated from northern blots. The testis *mia* cDNA also contained almost 600 bp of 5′ untranslated sequence containing multiple stop codons in all three reading frames upstream of the predicted initial methionine.

Fig. 3. Bar diagram comparisons and sequence alignments for the *nht*, *rye*, *mia* and *sa* predicted proteins. In bar diagrams: predicted histone fold domains are red; other conserved regions are blue. In multiple sequence alignments the colours are: yellow, hydrophobic; green, polar + Tyr; purple, charged; red, proline; orange, Phe. Red horizontal bars indicate alpha helices of the HFDs, shown above the alignment for the TAFs and below the alignment for the histones; blue or light blue horizontal bars indicate other conserved regions. (A,B) Structural comparisons of *nht* protein with TAF4 homologs from yeast to human. dTAF4: generally expressed *Drosophila* homolog of *nht.* Sequence alignments (B) show the histone fold domain (HFD) compared to histone H2a, the C-terminal CCTD region, and in between two additional regions with a conserved pattern of polar, charged and hydrophobic residues (dark-blue bars). (C,D) Structural comparisons of *rye* protein with TAF12 homologs from a range of metazoans. dTAF12: generally expressed *Drosophila* homolog of *rye.* Histone fold domain (HFD) also compared to histone H2b in (D). (E,F) Structural comparisons of *mia* protein with TAF6 homologs from yeast to human. dTAF6: generally expressed *Drosophila* homolog of *mia.* Sequence alignments (F) show the histone fold domain (HFD) compared to histone H4. The TAF6 domain is the extended region C-terminal to the HFD with a conserved pattern of polar, charged and hydrophobic residues. In B,D,F, residues with >15% of overall structure exposed (as predicted by DeepView software) in the hTAF4:hTAF12 or dTAF6 and dTAF9 heterodimers, based on their crystal structures (Werten et al., 2002; Xie et al., 1996) are marked ‘e’. Residues making heterodimeric bonds between hTAF4 and hTAF12, or between dTAF6 and dTAF9, analyzing all residues of opposing chains within 5 Å, are designated above the sequence alignment by the following symbols: 1 = binding to alpha-1 helix of heterodimer partner; 2 = binding to alpha-2 helix of heterodimer partner; 3 = binding to alpha-3 helix of heterodimer partner; i = forming an acid-base bond with heterodimeric partner. From the crystal structure of the nucleosome (Luger et al., 1997), regions of histones that form symmetrical tetrameric bonds are marked with an asterisk; regions of histones forming non-symmetrical tetrameric bonds (H2B-H4) are marked (8). (G,H) Comparisons of *sa* and TAF8 homologs from yeast to human. The histone fold domain predicted by PFAM family HMMer search is shown as a red bar, and the predicted alpha helices as green cylinders: secondary structure predictions show the helix-loop-helix-loop-helix pattern characteristic of histone fold domains. Significant region of homology just downstream of the predicted HFD marked with a light-blue bar.

Table 3

Gene pair	Number of aa in region aligned	% Identity	% Similarity	% Gap
A. Overall percent amino acid identity/similarity of full length testes TAFs with homologs in the multiple sequence alignments				
Nht:	259	–	–	–
TAF4	300	17	31	23
hTAF4	305	13	30	23
dH2A	123	6	13	61
Rye:	138	–	–	–
dTAF12	196	12	32	30
hTAF12	161	16	36	30
dH2B	122	5	22	36
Mia:	592	–	–	–
dTAF6	606	22	40	16
hTAF6	677	22	39	19
dH4	102	1	4	82
Sa:	275	–	–	–
dTAF8	328	14	26	32
hTAF8	310	15	28	32
B. Percentage identity/similarity of histone fold domain of testes TAFs to homologs in alignments				
Nht:	50	–	–	–
TAF4	45	14	38	10
hTAF4	45	10	46	10
dH2A	42	16	34	16
Rye:	67	–	–	–
dTAF12	68	29	63	1
hTAF12	68	32	63	1
dH2B	64	8	41	7
Mia:	62	–	–	–
dTAF6	63	19	47	1
hTAF6	63	20	46	1
dH4	63	9	39	1
Sa:	77	–	–	–
dTAF8	84	20	40	8
hTAF8	84	20	36	8
C. Pairwise percent identity/similarity of CCTD or TAF-6 conserved regions				
Nht CCTD:	31	–	–	–
dTAF4	28	48	77	9
hTAF4	28	32	67	9
Mia Cons Dom:	262			
dTAF6	262	39	64	1
hTAF6	262	38	65	1
Sa Cons Dom:	109			
dTAF8	100	18	32	15
hTAF8	100	20	38	15

The *mia* N-terminal region resembled corresponding regions of the generally expressed dTAF6 and TAF6 from other organisms, although conservation at the level of amino acid sequence was relatively low (Table 3). However, multiple sequence alignments using PFAM domain alignments and T-Coffee (see Materials and methods) revealed that amino acid residues 35-94 of the predicted Mia protein aligned with the histone fold domain of TAF6 and histone H4 with respect to the pattern of hydrophobic, charged and polar residues (Fig. 3F). Conservation of this pattern was strongest in the region extending from the middle of the alpha 2 helix through loop 2 and the alpha 3 helix of the HFD, based on the dTAF6:dTAF9 crystal structure (Xie et al., 1996). In the nucleosome, the regions of alpha 2, loop 2 and alpha 3 of histone H4 are involved both in dimerization, involving residues internal to the H3:H4 pair, and in forming non-symmetrical tetrameric bonds between H4 and H2B via exposed residues (Luger et al., 1997). The sequence of Mia as aligned in Fig. 3F threaded onto the crystal structure of dTAF6 in the dTAF6:dTAF9 heterodimer with an RMSD of 0.79 for aa33-aa94. In addition to the N-terminal predicted HFD, the Mia protein contained an extended central region (aa176-437 of Mia) with a pattern of hydrophobic, charged, polar and proline residues conserved in TAF6 proteins from yeast to humans. Embedded in this 'TAF6 domain' were several regions of significant amino acid sequence conservation (Table 3) between the predicted Mia protein, dTAF6 and TAF6 from other organisms (Fig. 3F).

The *sa* gene encodes a protein homologous to dTAF8, also known as *Prodos*. Sequencing of an 825 bp *sa* cDNA isolated by RT-PCR from testis mRNA revealed three protein coding exons separated by two introns (FBgn0037080). The cDNA had an in frame stop codon 84 base pairs upstream of the predicted initial methionine.

The N-terminal region of *sa* resembled the corresponding regions of the generally expressed dTAF8 (*Prodos*) and TAF8 from other organisms. Although conservation at the level of amino acid sequence was relatively low (Table 3), domain searches (PFAM) and multiple sequence alignments using T-Coffee (see Materials and methods) suggested that amino acid residues 4-70 of the predicted Sa protein contained a predicted histone fold domain, similar to TAF8 homologs from a variety of species with respect to the pattern of hydrophobic, charged and polar residues (Fig. 3H). Conservation of this pattern was strongest in the predicted alpha 2 helix region. Although no crystal structure was yet available for TAF8 homologs, secondary structure predictions by the Psipred program also suggested an α-helix – loop – α-helix – loop – α-helix pattern in this region of the predicted *sa* protein, consistent with a histone fold domain structure (Fig. 3H, green bars). In support of the prediction of an N-terminal histone fold domain in dTAF8, deletion of amino acids 1-39 from the generally expressed *Drosophila* TAF8 homolog *Prodos* disrupted binding between dTAF8 (*Prodos*) and its binding partner dTAF10B (Hernandez-Hernandez and Ferrus, 2001).

The predicted Sa protein also contained an extended central region (aa107-186) with significant amino acid sequence conservation (Table 3) and a pattern of hydrophobic, charged, polar and proline residues conserved in TAF8 homologs from yeast to humans (Fig. 3H). This region is embedded in a larger domain (aa78-186) listed in the pFam database as pFamB-10670, an unannotated domain characteristic of this family of TAFs (Fig. 3H, blue).

Two testis-specific TAF homologs with histone fold domains interact as testis-specific binding partners

Because TAF4 and TAF12 physically interact as binding partners through their histone fold domains (Gangloff et al., 2000; Werten et al., 2002; Yokomori et al., 1993), we tested whether the *Drosophila* testis-specific TAF4 and TAF12 homologs *nht* and *rye* also interact structurally, using a bacterial co-expression and GST pulldown assay (Fig. 4). In bacteria carrying a GST-*nht* fusion construct encoding full length Nht and an empty vector in place of a Rye-FLAG fusion construct, the GST-*nht* fusion protein was detected in total bacterial extracts (T) under inducing conditions, but was not soluble in the absence of Rye-FLAG fusion protein (Fig. 4, lane pair 1). A portion of Rye containing the histone fold region fused to a FLAG epitope was not stable and failed to accumulate when expressed in bacteria in the absence of Nht (Fig. 4, lane pair 2). However, when both the Nht and Rye fusion proteins were co-expressed in the same bacteria, both the GST-Nht fusion protein and the FLAG-tagged Rye HFD-containing fragment accumulated in the total bacterial extract (T) and were soluble. When the GST-Nht fusion protein was isolated from extracts from bacteria expressing both the Nht and Rye fusion proteins by binding to glutathione-Sepharose, the FLAG-Rye fusion bound and co-eluted with GST-Nht (Fig. 4, lane 3B).

The physical interaction between the Nht and Rye fusion proteins appeared to be specific for the testis-specific partners. The C-terminal region of the generally expressed *Drosophila* TAF4 homolog dTAF4, analogous to the full length Nht protein, interacted with dTAF12 in the same bacterial co-expression GST pull down assay, as previously shown (Gangloff et al., 2000): a FLAG epitope-tagged fusion protein

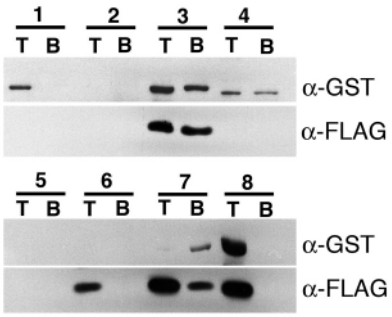


Fig. 4. The *nht* and *rye* proteins physically interact with each other but not with their generally expressed binding partner homologs TAF12 and TAF4 in a bacterial co-expression and GST pulldown assay. Western blots of total bacterial protein extracts (T) and elutions of glutathione-Sepharose purified protein (B) from strains expressing GST or FLAG-tagged fusion proteins. Blots were probed sequentially with anti-GST or anti-FLAG antibodies. Only proteins that were soluble in the bacterial extracts were able to be assayed for binding to glutathione-Sepharose. Lanes: 1, GST-Nht; 2, FLAG-Rye; 3, GST-Nht and FLAG-Rye; 4, GST-dTAF4 and FLAG-Rye; 5, GST-dTAF4; 6, FLAG-dTAF12; 7, GST-dTAF4 and FLAG-dTAF12; 8, GST-Nht and FLAG-dTAF12. In all cases, bacterial strains expressing a single fusion protein also carried the second vector without a fusion protein insert.

containing the histone fold region of dTAF12 was stable and co-purified with the GST-dTAF4 C-terminal domain fusion protein on glutathione-Sepharose (Fig. 4, lane 7B). However, under the same conditions, the C terminal (HFD containing) domain of the generally expressed dTAF4 fused to GST did not stabilize FLAG-tagged Rye when the two were co-expressed in bacteria (Fig. 4, lane pair 4). In similar assays, the GST-Nht fusion protein was not solubilized when co-expressed with the FLAG-dTAF12 HFD containing fragment, and so was not available to bind to and elute from the glutathione-Sepharose (Fig. 4, lane 8B).

Expression of the testis TAF mRNAs is mutually independent

To investigate whether the testis TAFs are transcribed independently at the onset of the primary spermatocyte program or whether some of the testis TAFs might regulate mRNA expression of the others, we assayed mRNA expression of *nht*, *can*, *mia*, *sa* and *rye* in spermatocytes from males mutant for *nht*, *can*, *mia* or *sa* by in situ hybridization to whole-mount testes. In all cases examined, mRNA for the testis TAFs accumulated in the various mutant spermatocytes (Fig. 5; Table 4). At times, the testis TAF transcripts appeared sharply at the boundary between spermatogonia and spermatocytes (Fig. 5), as in wild type. However, in some cases staining for the transcript appeared gradually in spermatocytes further from the testis apical tip. We could distinguish no clear pattern among these variations, which may be due in part to the probe or the degree to which spermatocytes are less crowded up into the testis apical third in the absence of differentiating spermatids in the mutants. Notably, for the alleles examined, *can* mRNA accumulated in *can* mutant spermatocytes. The same was true for *nht*, *sa*, and *mia* mRNA in *nht*, *sa*, or *mia* mutant spermatocytes, respectively (Fig. 5B,G; Table 4), indicating that transcription of any particular testis TAF did not depend on wild-type function of the respective protein itself. Transcripts for *can*, *mia*, *sa*, *nht* and *rye* also accumulated in spermatocytes mutant for *aly* (Fig. 5E,H), which has meiotic arrest and spermatid differentiation mutant phenotypes similar to the *can* class testis TAFs, although transcripts appeared to accumulate gradually rather than turn on abruptly in early spermatocytes. Accumulation of *aly* transcripts or protein was previously shown to be independent of wild-type function of the testis TAFs *can*, *mia* and *sa* (White-Cooper et al., 2000).

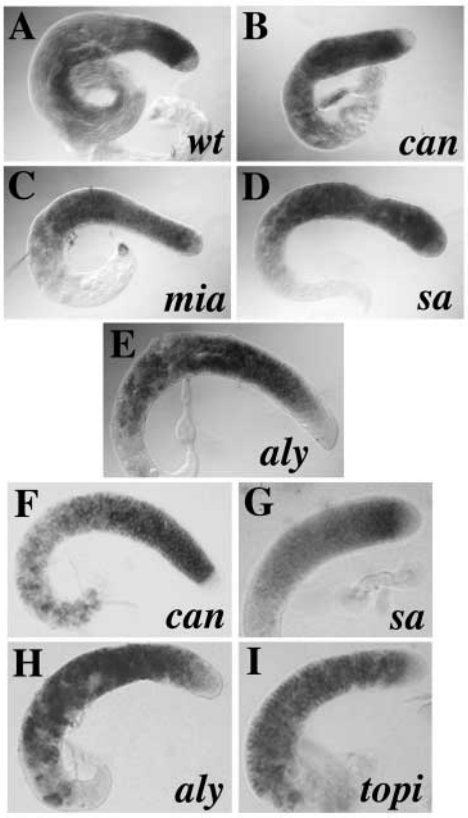


Fig. 5. Testis TAF transcripts accumulate in spermatocytes of testis TAF or *aly* mutant flies. In situ hybridization to testes using (A-E) *can* or (F-I) *sa* probes. (A) Wild type; (B) can^{3}; (C) mia^{1}; (D) *sa*; (E) *aly*; (F) can^{12}; (G) sa^{2}; (H) aly^{2}/aly^{5}; (I) *topi* testes.

Table 4. Testis TAFs are transcribed independently of each other and of *aly* class genes

	Genotype*				
Probe	*nht*	*can*	*mia*	*sa*	*aly*
nht	+†	+	+	+	+
can	+	+†	+	+	+
mia	+	+	+†	+	+
sa	+	+	+	+†	+
rye	+	+	+	+	+

*Allelic combinations used: can^{2}/can^{12}, $mia^{1}/mia^{z\text{-}3348}$, sa^{2}/sa^{2}, $nht^{z\text{-}5347}/nht^{z\text{-}5946}$, aly^{2}/aly^{5P}.

†For the particular allelic combinations used, mRNA was produced.

Discussion

Tissue-specific homologs of five different subunits of TFIID are expressed in differentiating male germ cells at the spermatocyte stage in *Drosophila.* In addition to *can*, which encodes a homolog of the WD40 repeat containing dTAF5 (Hiller et al., 2001) at least four histone fold domain containing TAFs have tissue-specific homologs: the *nht*, *mia*, *sa* and *rye* genes encode homologs of dTAF4, dTAF6, dTAF8 and dTAF12, respectively. Null mutations in *nht*, *can*, *sa* and *mia* all cause the same mutant phenotype and have similar effects on transcription in primary spermatocytes of several target genes involved in spermatid differentiation. In addition, the protein encoded by the *rye* testis TAF homolog interacts structurally and specifically with *nht*, suggesting that the five testis TAF homologs function together or in a pathway to regulate a gene selective transcription program for terminal differentiation in the male germline. Loss-of-function mutations in the *Drosophila* testis TAFs block spermatogenesis at the G_2/M transition of meiosis I, after spermatocyte growth, and the mutant testes have large numbers of differentiating and mature spermatocytes (Lin et al., 1996) (this study). Because transcription of spermatid differentiation genes occurs during the spermatocyte stage in *Drosophila*, the dramatic reduction in transcripts for target genes involved in spermatid differentiation in the mutants was not due to lack of the cell type in which the target genes are normally transcribed. Importantly, the requirement for the testis-specific TAF

homologs for normal levels of transcript accumulation of targets is gene selective; transcripts from many genes normally expressed in primary spermatocytes are expressed at normal levels in the testis TAF mutants (White-Cooper et al., 1998) (this study).

If the five testis TAFs identified to date act in a multi-subunit complex resembling TFIID, then the identity of the other subunits expected in a TFIID-like complex is a puzzle. The five TAF homologs we have identified may substitute for their more generally expressed homologs in a chimeric TFIID complex, where the other components are the generally expressed TAFs. Alternatively, the *Drosophila* testis TAFs may participate in a TFIID-like complex where the other components are highly diverged from their more generally expressed forms. We note that the *nht, can, mia, sa,* and *rye* predicted proteins are more diverged from their generally expressed homologs than the generally expressed *Drosophila* homologs are diverged from their human counterparts. Another possibility is that the testis TAF homologs act in a different type of protein complex, for example a HAT complex or with the Polycomb complex of chromatin modifying factors, to regulate the cell-type-specific expression of terminal differentiation genes required for spermatogenesis.

The testis TAF4 homolog *nht* lacked the long, glutamine-rich N terminus characteristic of both the generally expressed *Drosophila* homolog and TAF4 from mammals. This glutamine-rich N-terminal domain can interact structurally with certain transcriptional activator proteins in vitro and has been proposed to help TFIID mediate activated transcription, perhaps by tethering TFIID to transcriptional activators bound at enhancers (reviewed by Hochheimer and Tjian, 2003). If so, then the testis-specific homolog *nht* may render a possible testis-specific TFIID-like complex less sensitive to transcriptional activators that might normally interact through the glutamine rich N-terminal domain of dTAF4.

Northern blot analysis suggested low levels of alternate transcript forms of *mia* expressed in females and embryos. Alternate transcripts have also been described for TAF6 in human cells (Bell et al., 2001). However, analysis of null mutants, including an allele with an early stop codon in the *mia* open reading frame, suggested that wild-type function of *mia* is required for spermatogenesis but not for female fertility or embryonic development. It is possible that the *mia* protein is only expressed in spermatocytes, or that the generally expressed homolog dTAF6 can substitute for *mia* function in other tissues. Perhaps *mia* protein may be required to allow *nht:rye* to bind into a testis-specific TFIID or HAT-like complex, so that *mia* function is essential only where it is necessary to incorporate *nht:rye*. The HFD containing dTAF6 forms a heterodimer partner with dTAF9 (Xie et al., 1996). However, searches of the *Drosophila* genome have not yet revealed an obvious candidate for a second homolog of dTAF9 that might serve as a binding partner with *mia*.

The generally expressed TAF8 homolog *Prodos* binds specifically to dTAF10B in vitro and in yeast two-hybrid assays. The *Drosophila* genome encodes two homologs of TAF10. However, unlike the testis-specific TAF homologs we describe here, both dTAF10 and dTAF10B are expressed during embryogenesis, with some degree of tissue specificity (Georgieva et al., 2000). Preliminary tests in the bacterial co-expression assay did not reveal interaction between the testis-specific dTAF8 homolog *sa* and either dTAF10 or dTAF10B (X.C., unpublished data). Searches of the *Drosophila* genome have not yet revealed an obvious candidate for an additional dTAF10 homolog that might serve as an HFD heterodimer partner with *sa*.

Tissue-specific GTFs and the male germline

Several tissue-specific TAFs and other GTF component homologs have been found to be expressed in the testis in differentiating male germ cells in mammals as well as in *Drosophila*. Altering the composition of the general transcription machinery may be particularly important for gene expression in the male germline. Differentiation of male gametes in both *Drosophila* and mammals depends on a robust germline-specific transcriptional program (Fuller, 1993; Goldberg, 1996; Hecht, 1993). Expression of many genes required for spermatid differentiation takes place in spermatocytes in *Drosophila* and in spermatocytes and/or early round spermatids in mammals. Many genes that are transcribed in somatic cells at other stages of development are expressed in male germ cells from testis-specific promoters. In addition, a number of generally expressed genes in *Drosophila* have homologs that are only or mainly expressed in the testis. In several cases the cis-acting regulatory sequences that drive expression of the testis-specific transcripts have been shown to be contained within short regions positioned near the start of transcription. Humans express a testis-enriched subunit of TFIIA and a testis-specific TAF1 homolog (Ozer et al., 2000; Upadhyaya et al., 1999; Wang and Page, 2002) while a testis-specific homolog of TAF7 has been identified in mouse (Pointud et al., 2003). In addition, wild-type function of the TATA-binding protein homolog TRF2 is required in mouse to produce mature sperm (Martianov et al., 2001; Zhang et al., 2001). Both our studies of the testis TAFs of *Drosophila* and studies of knockout mutant mice lacking mTRF2 function indicate that these testis-specific homologs of GTF components are required for normal transcription and terminal differentiation in male germ cells. It is possible that chromatin may be in a different condensation state in spermatocytes and early round haploid spermatids than in many somatic cells and so may require specialized forms of the general transcription machinery to recognize or access testis-specific promoters within an altered chromatin landscape.

One striking finding was the tissue specificity of the structural interactions between dTAF4:dTAF12 compared with the *nht:rye* proteins. Examination of the virtual structure of the histone fold domains of *nht* and *rye* threaded onto the crystal structure of the hTAF4:hTAF12 HFD heterodimer did not reveal any obvious single reason for the specificity of the dimer partners observed in biochemical assays, suggesting that the specificity of the binding partner interaction may be the result of an additive effect of a number of residue interactions across the HFD. One notable difference between the predicted structures of dTAF4:dTAF12 compared with *nht:rye*, based on threading onto the hTAF4:hTAF12 HFD crystal structure, appeared in the region where loop1 of TAF4 interacts with loop2 of TAF12. Loop 1 of hTAF4 and loop 2 of hTAF12 both have a short region of beta sheet, which interacts in parallel to anchor the end of the α2 helix interaction in the hTAF4:hTAF12 crystal structure (Werten et al., 2002). The dTAF4:dTAF12 sequences fit well across this region, yielding

predicted interacting beta sheets based on the threading algorithm used (SwissModel). However, the *nht:rye* predicted structures did not fit this region well. Both *nht* loop 1 and *rye* loop 2 lacked a predicted short beta sheet in the virtual heterodimer formed by threading onto the hTAF4:hTAF12 crystal structure. When we took the virtual structures of dTAF4 and *rye* predicted by threading and placed them in the heterodimer positions, an acidic clash between Asp732 of the dTAF4 HFD and Asp117 of *rye* was created in the region where loop1 of TAF4 would contact loop 2 of *rye*. A second notable difference in the predicted heterodimers of dTAF4:dTAF12 and *nht:rye* was that the *nht* His-75:*rye* Ser-104 interaction contained a hydrogen bond. This interaction would be expected to be much stronger than the weak Ala-Val interaction at the corresponding position in TAF4:TAF12. This interaction would also be abrogated in an *nht*:TAF12 or TAF4:*rye* heterodimer. However, the alpha-1 helix of TAF4, corresponding to the region where *nht* His-75 is located, was not required for TAF4-TAF12 dimerization in yeast TAF4 mutant rescue experiments (Thuault et al., 2002). The dTAF4:dTAF12 structure predicted by threading on to the hTAF4:hTAF12 crystal structure is likely to be relatively reliable (RMSD=0.14A). However, the amino acid sequences of the predicted HFD motifs of *nht* and *rye* were considerably diverged from the corresponding regions of both the human and the generally expressed *Drosophila* TAF4 and TAF12 proteins (Table 3). As a result, the virtual structure of the *nht:rye* HFD heterodimer, calculated from threading, is much less reliable (RMSD=0.74A) and may underestimate the divergence of the protein structures from the heterodimer between the generally expressed homologs. Since the *nht:rye* heterodimer can be stably expressed in bacteria (Fig. 4) the best way to compare the structures and probe the molecular basis of the specificity between the generally expressed and the testis-specific predicted heterodimers may be through solving the crystal structure of the testis-specific *nht:rye* complex.

We thank Caroline Chi and Anna Ballew for assistance on the cloning of *nht*, J. Root and M. Ashburner for deficiencies in the *nht* region, D. Lindsley for the original allele of *sa*, J. Hachstein for the original allele of *mia*, B. Wakimoto, D. Lindsley and C. Zuker for male sterile alleles of *mia* and *nht*, S. Hyashi for flies bearing the *nht* rescue construct, the Berkeley Drosophila Genome Project for genome sequences and the testis AT cDNA library, C. O'Connor for the cosmid containing the *mia* region, S. Fribourg and D. Moras for the bacterial co-expression vectors, and Y.-G. Gangloff, and M. Levitt for helpful discussions. This work was supported by an NICHD Reproductive Biology Training Grant, a Stanford Medical School Dean's Fellowship, and a Laylor Foundation Fellowship to M.H., a Genome Science Training Grant number 2T32 HG00044 to M.J.P., NIH training grant no. 2T52GM07790 to S.M., NIGMS MSTP grant no. GM07365 to T.-Y. Lin, and NIGMS grant no. R01-GM61986 to M.T.F.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.

Albright, S. R. and Tjian, R. (2000). TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**, 1-13.

Aoyagi, N. and Wassarman, D. A. (2000). Genes encoding Drosophila melanogaster RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. *J. Cell Biol.* **150**, F45-50.

Ayyar, S., Jiang, J., Collu, A., White-Cooper, H. and White, R. A. (2003). Drosophila TGIF is essential for developmentally regulated transcription in spermatogenesis. *Development* **130**, 2841-2852.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res.* **30**, 276-280.

Bell, B., Scheer, E. and Tora, L. (2001). Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function. *Mol. Cell* **8**, 591-600.

Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge, UK: Cambridge University Press.

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E. and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* **293**, 2084-2087.

Fribourg, S., Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A. and Moras, D. (2001). Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. *J. Mol. Biol.* **306**, 363-373.

Fuller, M. T. (1993). Spermatogenesis. In *The Development of Drosophila melanogaster* (ed. M. Bate and A. Martinez-Arias), pp. 71-147. New York: Cold Spring Harbor Laboratory Press.

Gangloff, Y. G., Werten, S., Romier, C., Carre, L., Poch, O., Moras, D. and Davidson, I. (2000). The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs. *Mol. Cell Biol.* **20**, 340-351.

Georgieva, S., Kirschner, D. B., Jagla, T., Nabirochkina, E., Hanke, S., Schenkel, H., de Lorenzo, C., Sinha, P., Jagla, K., Mechler, B. et al. (2000). Two novel Drosophila TAF(II)s have homology with human TAF(II)30 and are differentially regulated during development. *Mol. Cell Biol.* **20**, 1639-1648.

Goldberg, E. (1996). Transcriptional regulatory strategies in male germ cells. *J. Andrology* **17**, 628-632.

Guan, K. L. and Dixon, J. E. (1991). Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* **192**, 262-267.

Guex, N. and Peitsch, M. C. (1997). *Electrophoresis* **18**, 2714-2723.

Hecht, N. B. (1993). Gene expression during male germ cell development. In *Cell and MolecularBiology of the Testis* (ed. C. Desjardins and L. L. Ewing), pp. 400-432. New York: Oxford University Press.

Hernandez-Hernandez, A. and Ferrus, A. (2001). Prodos is a conserved transcriptional regulator that interacts with dTAF(II)16 in Drosophila melanogaster. *Mol. Cell Biol.* **21**, 614-623.

Hiller, M. A., Lin, T. Y., Wood, C. and Fuller, M. T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.* **15**, 1021-1030.

Hochheimer, A. and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev.* **17**, 1309-1320.

Jiang, J. and White-Cooper, H. (2003). Transcriptional activation in Drosophila spermatogenesis involves the mutually dependent function of aly and a novel meiotic arrest gene cookie monster. *Development* **130**, 563-573.

Lemon, B. and Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14**, 2551-2569.

Lin, T. Y., Viswanathan, S., Wood, C., Wilson, P. G., Wolf, N. and Fuller, M. T. (1996). Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* **122**, 1331-1341.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260.

Martianov, I., Fimia, G. M., Dierich, A., Parvinen, M., Sassone-Corsi, P. and Davidson, I. (2001). Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol Cell* **7**, 509-515.

Nicholas, K. B., Nicholas, H. B. J. and Deerfield, D. W. I. (1997). *EMBNEW. NEWS* **4**, 14.

Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217.

Orphanides, G., Lagrange, T. and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657-2683.

Ozer, J., Moore, P. A. and Lieberman, P. M. (2000). A testis-specific transcription factor IIA (TFIIAtau) stimulates TATA-binding protein-DNA binding and transcription activation. *J. Biol. Chem.* **275**, 122-128.

Perezgasga, L. C., Jiang, J., Bolival, J., B., Hiller, M., Fuller, M. T. and White-Cooper, H. (2004). Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn finger protein *matotopetli*. *Development* **131**, 1691-1702.

Pointud, J. C., Mengus, G., Brancorsini, S., Monaco, L., Parvinen, M., Sassone-Corsi, P. and Davidson, I. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. *J. Cell Sci.* **116**, 1847-1858.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327-335.

Rubin, G. M. and Spradling, A. C. (1983). Vectors for P element-mediated transformation in *Drosophila*. *Nucleic Acids Research* **11**, 6341-6351.

Sandler, L., Lindsley, D. L., Nicoletti, B. and Trippa, G. (1968). Mutants affecting meiosis in natural populations of *Drosophila melanogaster*. *Genetics* **60**, 525-558.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.

Thuault, S., Gangloff, Y. G., Kirchner, J., Sanders, S., Werten, S., Romier, C., Weil, P. A. and Davidson, I. (2002). Functional analysis of the TFIID-specific yeast TAF4 (yTAF(II)48) reveals an unexpected organization of its histone-fold domain. *J. Biol. Chem.* **277**, 45510-45517.

Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* **16**, 673-675.

Tower, J., Karpen, G. H., Craig, N. and Spradling, A. C. (1993). Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. *Genetics* **133**, 347-359.

Upadhyaya, A. B., Lee, S. H. and DeJong, J. (1999). Identification of a general transcription factor TFIIAalpha/beta homolog selectively expressed in testis. *J. Biol. Chem.* **274**, 18040-18048.

Veenstra, G. J. and Wolffe, A. P. (2001). Gene-selective developmental roles of general transcription factors. *Trends Biochem. Sci.* **26**, 665-671.

Verrijzer, C. P. (2001). Transcription factor IID–not so basal after all. *Science* **293**, 2010-2011.

Wakimoto, B. T., Lindsley, D. L. and Herrera, C. (2004). Toward a Comprehensive Genetic Analysis of Male Fertility in Drosophila melanogaster. *Genetics* **167**, 207-216.

Wang, P. J. and Page, D. C. (2002). Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum. Mol. Genet.* **11**, 2341-2346.

Wang, Z. and Mann, R. S. (2003). Requirement for two nearly identical TGIF-related homeobox genes in Drosophila spermatogenesis. *Development* **130**, 2853-2865.

Werten, S., Mitschler, A., Romier, C., Gangloff, Y. G., Thuault, S., Davidson, I. and Moras, D. (2002). Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J. Biol. Chem.* **277**, 45502-45509.

White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M. T. (2000). Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* **127**, 5463-5473.

White-Cooper, H., Schafer, M. A., Alphey, L. S. and Fuller, M. T. (1998). Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development* **125**, 125-134.

Xie, X., Kokubo, T., Cohen, S. L., Mirza, U. A., Hoffmann, A., Chait, B. T., Roeder, R. G., Nakatani, Y. and Burley, S. K. (1996). Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316-322.

Yagi, Y. and Hayashi, S. (1997). Role of the Drosophila EGF receptor in determination of the dorsoventral domains of escargot expression during primary neurogenesis. *Genes Cells* **2**, 41-53.

Yokomori, K., Chen, J. L., Admon, A., Zhou, S. and Tjian, R. (1993). Molecular cloning and characterization of dTAFII30 alpha and dTAFII30 beta: two small subunits of Drosophila TFIID. *Genes Dev.* **7**, 2587-2597.

Zhang, D., Penttila, T. L., Morris, P. L., Teichmann, M. and Roeder, R. G. (2001). Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science* **292**, 1153-1155.