Analysis

Genes Encoding Drosophila melanogaster RNA Polymerase II General Transcription Factors: Diversity in TFIIA and TFIID Components Contributes to Gene-specific Transcriptional Regulation

Norikazu Aoyagi and David A. Wassarman

National Institutes of Health, National Institute of Child Health and Human Development, Cell Biology and Metabolism Branch, Bethesda, Maryland 20892

How have the factors required for transcription initiation (TFIIA, TFIBM, TFID, TFIE, TFIF, and RNA polymerase II [pol II]) evolved to accommodate the elaborate transcriptional programs required for growth, differentiation, and development of multicellular organisms? Here we present analysis of the recently completed Drosophila melanogaster genome sequence, as well as those of Caenorhabditis elegans, Saccharomyces cerevisiae, and humans, that sheds light on this well studied question in eukaryotic biology. All four organisms encode single isoforms of RNA pol II, TFIBM, TFID, TFIE, and TFIIH components, but multiple, sequence-related isoforms of TFIIA components (Fig. 1; Albright and Tjian, 2000). In addition, Drosophila and humans encode multiple isoforms of TFIIA components (U padhyaya et al., 1999; Ozer et al., 2000). Current evidence indicates that tissue- and cell type-specific transcription is directed by differentially expressed TFIIA and possibly TFID isoforms (Zeidler et al., 1996; Albright and Tjian, 2000). Thus, in accord with experimental data, this analysis points to TFIIA and TFID as the factors that help generate the broad transcriptional repertoire of multicellular organisms. The identification of the complete set of TFIIA and TFID components in a genetically and biochemically tractable organism like Drosophila is an important step toward understanding the mechanisms governing developmentally regulated transcription not only in Drosophila but also in humans.

The Biology of Transcription Initiation

Biochemical fractionation of Drosophila embryos, human cells, and yeast cells has defined a set of multiprotein complexes termed general transcription factors (GTFs; TFIIA, TFIBM, TFID, TFIE, TFIF, and TFIIH) required for mRNA transcription initiation in vitro (Orphanides et al., 1996; Hampsey, 1998). Transcription is initiated by recognition of core promoter elements by TFIIID and sequential or concerted assembly of the other GTFs and RNA pol II to form the preinitiation complex (PIC). Although GTFs play essential roles during transcription initiation, it is the factors that regulate the ability of the GTFs to assemble and stably bind a core promoter that are probably major determinants of gene-specific transcription levels. For example, activators and coactivators are thought to stimulate transcription by recruiting GTFs to a promoter, thereby accelerating PIC assembly.

The GTF TFIIID is composed of TATA-binding protein (TBP) and coactivator subunits termed TBP-associated factors (TAFIIIs; Burley and Roeder, 1996; Green, 2000; Albright and Tjian, 2000). TAFIIIs not only function as “conventional” coactivators by serving as physical links between DNA-binding activator proteins and the PIC but also possess enzymatic or promoter recognition activities that presumably enhance the efficiency of PIC assembly. TFIIA has also been described as a coactivator and displays a number of TAFIIIs-like properties: it binds to TBP and TAFIIIs; it interacts with specific transcriptional activators; it is generally required for activated transcription in vitro; and it contributes to promoter selectivity (Orphanides et al., 1996; Hampsey, 1998).

TAFII, TBP, and TFIIA Components Mediate Gene-specific Transcription

Inactivation of individual TAFIIIs in Drosophila, mammalian, and yeast cells has demonstrated that TAFIIIs are not required for the transcription of all RNA pol II genes, and in fact there is great variation in regard to the identity and number of gene targets for individual TAFIIIs (Green, 2000; Albright and Tjian, 2000). Furthermore, different domains within a single TAFII can play gene-specific roles in transcription (O’Brien and Tjian, 2000). The isolation of a
Figure 1. RNA pol II and GTF-encoding subunit genes from D. melanogaster, C. elegans, humans, and yeast are grouped by protein complex. Additional information about these genes can be accessed using the indicated gene name or identification number at web sites for the Drosophila (http://www.fruitfly.org/annot/), C. elegans (http://www.wormbase.org/), yeast (http://www.proteome.com/databases/index.html), or human genomes (http://www.ncbi.nlm.nih.gov/). No Gadfly identification has been assigned for Rpb12 because it was not identified by gene prediction programs and no expressed sequence tags have been isolated. However, searches using the human and yeast Rpb12 homologues match translated genomic sequence from EMBL/GenBank/DDJB accession No. A19709, and further analyses of sequences surrounding this match reveal additional amino acid similarity that spans 3 exons (our unpublished observation). Each row contains homologous genes from each of the four organisms. An asterisk indicates that the gene is alternatively spliced. ND indicates information that has not been determined. The following identification numbers correlate to predicted mRNA s that in addition to encoding the indicated protein may also encode another protein, presumably due to a gene prediction error: Drosophila CG 7150, CG 6572, and C. elegans CG6572, CG7150, and CG6572, and yeast (http://www.proteome.com/databases/index.html). A search program that identifies protein motifs [http://www.isrec.isb-sib.ch/software/PFSCAN_form.html] and visual comparison of sequences was used to find the histone fold motifs in TFIIs and the transcription factors indicated in the text. Y. lactis, TAF II48, and TAF II65 are recently described components of TFIID (Sanders and Weil, 2000; Matangkasombat et al., 2000; Reese et al., 2000). Y. lactis TAFII145 and BDF1 display functional and sequence similarity to the NH2 and COOH termini, respectively, of human TAFII250 and therefore are placed in the same box (Matangkasombat et al., 2000). Bdf1 and Bdf2 display sequence similarity but only Bdf1 has been demonstrated to associate with TFIID. TAFI130/A NC1 is also a component of NuA3, TFIIF, and SWI/SNF complexes (Joh et al., 2000).
human B cell–specific isoform of TFIIA130 (TFIIA105) raised the possibility that substoichiometric subunits of TFIIID mediate tissue- or cell type–specific transcription and that additional components of TFIIID may have escaped detection because of their low abundance (Dikstein et al., 1996; Y amit-H ezi et al., 2000). These possibilities have been born out in Drosophila where isoforms of TAFF1110 and TAFF180 (No hitter [Nht] and Cannonball [Can], respectively) are expressed exclusively in testis and regulate transcription of a subset of genes required for spermatogenesis, and isoforms of TBP (TBP-related factors [TRF1 and TRF2]) are expressed in a tissue-specific manner and bind different genes in salivary gland cells (Hansen et al., 1997; Rabenstein et al., 1999; Hiller, M., T.-Y. Lin, and M. Fuller, personal communication). Similarly, analysis of the human TFIIA-L isoform ALF (TFIIA α/β-like factor) reveals that its expression is restricted to the testis; however, it remains to be determined if it is used for the transcription of testis-specific genes (U padhyaya et al., 1999; O zer et al., 2000). In Drosophila, TFIIA-S is expressed in a dynamic pattern during eye development and is transiently upregulated in photoreceptor precursor cells before their fate is determined (Zeidler et al., 1996). Therefore, the role of TFIIA and TFIIID in transcription initiation is governed by the expression patterns and activities of their varied components.

Finally, it is critical to note that analysis of the function of TFIIIS is complicated by the fact that they are components of at least two other complexes that lack TBP, p300/CBP-associated factor (PCAF) and TBP-free TFIIID containing complex (TFTC) (Struhl and M oqtaderi, 1998; Bell and T ora, 1999). The human PCAF histone acetyltransferase (HAT) complex contains three TAFII60s that are shared with TFIIID (TFIIA31/32, TAFII20/15, and TAFII30) and three TAFII60 isoforms (PCAF-associated factor 65β [PA F65β], PA F65α, and SPT3) related to TA FII1100, TAFII70/80, and TAFII18, respectively (Birck et al., 1998; O gy rzko et al., 1998). Yeast possess an analogous complex, S pr-A da-Gcn5-ace tyltransferase (SA GA), containing TFIIID TAFII60s and the Gcn5 HAT, and Drosophila may also, as it contains a Gcn5/PCAF homologue that interacts with TAFII24 (Smith et al., 1998; Brown et al., 2000; G eorgieva et al., 2000).

The Genomics of Transcription Initiation

Searches of the completed Drosophila, C. elegans, and yeast genomes and the partial human genome for sequence homologues of biochemically identified components of the general transcription machinery have led to the following conclusions. First, all of the components of RNA pol II, TFII B, TFII E, TFII F, and TFII H are encoded by single copy genes in Drosophila, C. elegans, and yeast (Fig. 1 A). Second, multiple isoforms of TFII D components are encoded in Drosophila, C. elegans, humans, and yeast, and multiple isoforms of TFIIA components are encoded in Drosophila and humans (Fig. 1 B). Third, each organism encodes isoforms of different sets of TFIIA and TFII D components, some which are unique to a particular organism.

Sequence comparisons uncovered Drosophila homologues of TFII S previously identified in yeast or humans by biochemical means but which had not been described in Drosophila (yeast TAFII67/human TAFII19, yeast TAFII30/ human ENL/A F-9, and yeast TAFII19/human TAF II18; G reen, 2000). Thus, all TAFII S present in both yeast and humans are present in Drosophila, as well as C. elegans. In contrast, yeast TAFII47 and TAFII65 are absent from Drosophila, C. elegans, and apparently from humans, suggesting that these TAFII S perform a yeast-specific role, such as serving as coactivators for DNA-binding activators that are not present in metazoans. Finally, there are TAFII S present in Drosophila, C. elegans, and humans that are absent from yeast (human TAFII68/Drosophila Cabeza and multiple TAFII isoforms). In addition to Can and Nht, there are alternatively spliced forms of TAFII30α, two genes (TAFII24 and TAFII16) that encode Drosophila homologues of human TAFII30, and TAFII60 and TAFII30 isoforms (TAFII60-2 and TAFII30-2, respectively) (K okubo et al., 1994; G eorgieva et al., 2000). TFIIA-S and TFIIA-L are the only other GTF components in Drosophila and humans, respectively, that are expressed in multiple isoforms (U padhyaya et al., 1999; O zer et al., 2000). The fact that these proteins are unique to multicellular organisms suggests that they play cell-specific roles.

A number of TAFII S contain a common structural motif called the histone fold that was originally shown to drive folding and association of each of the core histones (H2A, H2B, H3, and H4) and subsequently shown to play a similar role in association of TAFII S (Xie et al., 1996; Wolfe, 1998). TAFII α pairs, such as Drosophila TAFII40 and TAFII60, form heterotetramers, analogous to H3 and H4, and numerous other TAFII α-TAFII α and TAFII α-nonTAFII α interactions have been shown to involve histone fold motifs (G angeloff et al., 2000). The demonstrated histone fold interaction of human TAFII135 and TAFII20, predicts that Drosophila isoforms of these proteins, Nht and TAFII30α-2, respectively, may heterodimerize and hints at the existence of a human TAFII20 isoformal that would heterodimerize with the TAFII 135 isoformal, TAFII105. B cell–specific expression of the hypothetical TAFII20 isoformal may explain why TAFII105 associates with TFIIID in B cells but not in other cell types (Dikstein et al., 1996).

In addition to the TAFII S indicated in Fig. 1 B, other Drosophila transcription factors contain histone fold motifs: Prodos (Drosophila genome project Gadfly accession number CG728), NF-Y B-like (CG10477), NF-Y C-like (CG3075, CG11301), CH RA C-14 (CG13399), CH RA C-16 (CG15736), Dr1 (CG4185), NC2x (CG10318), and BIP2 (CG2009). It is interesting to speculate that these factors may be unidentified TAFII α components of TFIIID or binding partners for known TAFII S in complexes that lack TBP.

Putting It All Together

A synthesis of eukaryotic genomes has defined sets of proteins that are similar in sequence to known components of TFII A and TFII D. Since known components of TFII A and TFII D have been shown to playa key roles in developmentally regulated transcription, it is exciting to speculate that the newly identified genes will play similar roles and that TFII A and TFII D components have evolved to support tissue- or cell type–specific transcriptional requirements of individual eukaryotic organisms.
The challenge now is to determine if TAF IIs that have been identified on the basis of their sequence are components of TBP-containing complexes or other TAF IIs-containing complexes, whether TAF IIs and TFIIA isoforms are differentially expressed during development, and how differentially expressed TBP, TAF IIs, and TFIIF isoforms function in concert with the ubiquitously expressed form of TFIIF and TFIID to regulate gene expression. The subunit composition of human PCAF complex leads to the prediction that Drosophila TAF II60-2 and Can and C. elegans Y37E11A.6.c are components of PCAF/SA GA and not TFIID. On the other hand, protein isoforms that are unique to a particular organism, such as Drosophila TAF II30 and TAF II30a-2 and C. elegans F54F7.1 and K10D 3.3, may be tissue- or cell type-specific components of TFIID and not PCAF/SA GA.

Drosophila may be the most appropriate organism for these studies since the biochemical activities of these factors can be determined using established TFIIF and TFIID purification schemes and in vitro transcription systems, and developmental requirements for these factors can be determined using existing mutants or mutants generated by traditional mutagenesis schemes, P-element insertion, RNA interference (RNAi), or homologous recombination (Kennerdell and Carthew, 1998; Rorth et al., 1998; Spradling et al., 1999; Rong and Golic, 2000).

In terms of the RNA pol II transcriptional machinery, this review has covered only the tip of the iceberg. Detailed analysis of Drosophila genes encoding DNA-binding transcription factors, coactivators, corepressors, chromatin remodeling factors, and other trans-acting regulators of transcription remains to be tackled. However, completion of the Drosophila genome sequence has set the stage for biochemical, molecular, and genetic studies in Drosophila that should lead to advances in our understanding of developmentally regulated RNA pol II transcription.

In addition to being able to identify new components of the transcription machinery, the Drosophila genome project has provided several valuable tools for studying RNA pol II transcription. First, it has led to the identification of fly stocks containing P-element insertions that disrupt GTF genes, providing the opportunity to investigate developmental and possibly mechanistic roles for the encoded factors (Rorth et al., 1998; Spradling et al., 1999). Second, sequencing of full-length expressed sequence tags (i.e., cDNA s) has helped define RNA pol II transcription start sites that may lead to the identification of novel core promoter elements or provide insight into how different combinations of core promoter elements contribute to transcription initiation. The recent description of a TAF-rich sequence (TC-box) that is specifically bound by Drosophila TRF1 and the identification of isoforms (i.e., TAF II60-2) of known TAF II60 (i.e., TAF II60) that recognize core promoter elements hints at the existence of additional core promoter elements (Burke et al., 1998; Holmes and Tjian, 2000). Finally, the description of the ~13,600 Drosophila genes allows for construction of DNA microarrays (i.e., gene chips) that can be used to identify gene targets for individual components of the transcription machinery (A dams et al., 2000).

We thank M. Hiller, T.-Y. Lin, and M. Fuller for providing unpublished data on Can and Nht; Celera Genomics and the Berkeley Drosophila Genome Project for allowing us to search the Drosophila genome sequence before publication; Y. Le for assistance performing sequence searches; and R. Kamakaka, L. Pile, P. Wade, and K. Wasserman for providing comments on the manuscript.

This work was supported by the Intramural Program in the National Institute of Child Health and Human Development.

Submitted: 7 Apr 2000
Revised: 22 Jun 2000
Accepted: 22 Jun 2000

References

A dams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. et al. 2000. The genome sequence of Drosophila melanogaster. Science. 287:2185–2195.

Bell, D., and L. Tora. 1999. Regulation of gene expression by multiple forms of TFIID and other TAF IIs-containing complexes. Exp. Cell Res. 246:11–19.

Birck, C., O. Poch, C. Romier, M. Ruff, G. Mengs, A.-C. Lavigne, I. Davidson, and D. M. 1998. TAF II28 and TAF II26 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT 3 family. Cell. 94:239–249.

Brown, C. E., T. Lechner, L. Howe, and J. L. Workman. 2000. The many hats of transcriptional coactivators. Trends Biochem. Sci. 25:15–19.

Burke, T. W., P. J. Wilky, A. K. Kuchta, E. F. Butler, and J. T. K. K. 1998. A conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harbor Symp. Quant. Biol. 63: 75–82.

Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). A. Rev. Biochem. 65:769–799.

Dikstein, R., S. Zhou, and R. Tjian. 1996. Human TAF II 105 is a cell type-specific TFIID subunit related to H TAF II 130. Cell. 87:137–146.

Gangloff, Y.-G., S. Werten, G. Romier, L. Carre, O. Poch, D. M., and I. Davidson. 2000. The human TFIID components TAF II 135 and TAF II 20 and yeast SAGA components ADA 1 and TAF II 68 homodimerize to form histone-like pairs. Mol. Cell Biol. 20:340–351.

Georgieva, S., D. B. Kirshner, T. Jagla, E. Nabirochkina, S. Hanke, H. Schenkel, C. de Lorenzo, P. Sinha, K. Jagla, B. Mchelth, and L. Tora. 2000. Two novel Drosophila TAF II s, Howe, and TFIIA are differentially regulated during development. Mol. Cell Biol. 20:1639–1648.

Green, M. R. 2000. TBP-associated factors (TAF II s): multiple, selective transcriptional mediators in common complexes. Trends Biochem. Sci. 25:59–63.

Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcription machinery. Microbiol. Mol. Biol. Rev. 62:465–503.

Hansen, S. K., S. Takada, R. H. Jacobson, J. T. Lis, and R. Tjian. 1997. Transcription properties of a cell type-specific TATA-binding protein-trancript, TRF. Cell. 91: 71–83.

Holmes, M. C., and R. Tjian. 2000. Promoter-selective properties of the TBP-related factor TR F1. Science. 288:867–870.

John, S., L. Howe, S. T. Tafrow, P. A. Grant, R. Sternglanz, and R. Workman. 2000. TheSomething about the activity of the TAF II30 protein, TAF II30, is the catalytic subunit of NuA 3, a yTA FII130-containing HAT complex that interacts with the sp16 subunit of the yeast CP (Cdc68p/Pol3-1)-FACT complex. Genes Dev. 14:1196–1208.

Kennerdell, J. R., and R. W. Carthew. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell. 95:1017–1026.

Kekubo, T., D.-W. Gong, J. C. Woolott, M. Horikoshi, R. G. Roeder, and Y. Nakatanai. 1994. Molecular cloning of Drosophila TFIID subunits. Nature. 367:484–487.

Matangkasombut, O., R. M. Buratowski, N. W. Swilling, and S. Buratowski. 2000. Bromodomain factor 1 corresponds to a missing piece of yeast TFIIF. Gene. 25:14:951–962.

O’Brien, T., and R. Tjian. 2000. Different functional domains of TFII F, 250 modulate expression of distinct subsets of mammalian genes. Proc. Natl. Acad. Sci. USA. 97:2456–2461.

Ogzysko, V. V., T. Kotani, X. Zhang, R. L. Schiltz, T. Howard, X.-j. Yang, B. H. Howard, J. Qin, and Y. Nakatanai. 1998. Histone-like TAFs within the PCAF histone acetylase complex. Cell. 94:35–44.

Orphanides, G., T. Largrange, and D. R. 1996. The general transcription factors of RNA polymerase II. Genes Dev. 10:2657–2683.

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

Rabenstein, M. D., S. Zhou, J. T. Lis, and R. Tjian. 1999. TATA box-binding protein (TBP)-related factor 2 (TR F2), a third member of the TBP family. Proc. Natl. Acad. Sci. USA. 96:4791–4796.

Rong, Y. S., Z. Zhang, and H. Kuped. 2000. Identification of a yeast transcription factor II D subunit, TSG 28FA F88. J. Biol. Chem. 275:17391–17398.

Rong, Y. S., and K. G. Golic. 2000. Gene targeting by homologous recombination in Drosophila. Science. 288:2013–2016.
Röth, P., K. Szabo, A. Bailey, T. Laverty, J. Rehm, G.M. Rubin, K. Weigmann, M., M. Milan, V. Benes, W. Arisoge, and S.M. Cohen. 1998. Systematic gain-of-function genetics in Drosophila. Development. 125:1049-1057.

Sanders, S.L., and P.A. Weil. 2000. Identification of two novel TAF subunits of the yeast Saccharomyces cerevisiae TFIID complex. J. Biol. Chem. 275:13895-13900.

Smith, E.R., J.M. Belote, R.L. Schiltz, X.J. Yang, P.A. Moore, S.L. Berger, Y. Nakatani, and C.D. Allis. 1998. Cloning of Drosophila GCN5: conserved features among metazoan GCN5 family members. Nucleic Acids Res. 26:2948-2954.

Spradling, A.C., D. Stern, A. Beaton, E.J. Rhem, T. Laverty, N. Mozden, S. Misra, and G.M. Rubin. 1999. The Berkeley Drosophila Genome Project gene disruption project: single P-element insertions mutating 25% of vital Drosophila genes. Genetics. 153:135–177.

Struhl, K., and Z. Moqtaderi. 1998. The TAFs in the HAT. Cell. 94:1–4.

Upadhaya, A.B., S.H. Lee, and J. Dejong. 1999. Identification of a general transcription factor TFIIFα homologue selectively expressed in testis. J. Biol. Chem. 274:18040-18048.

Wolffe, A.P. 1998. Chromatin structure & function. Academic Press, New York. 7-157.

Xie, X., T. Kokubo, S. Cohen, U. Mirza, A. Hoffman, B. Chait, R. Roeder, Y. Nakatani, and S. Burley. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. Nature. 380:316-322.

Yamit-Hezi, A., S. Nir, O. Wolstein, and R. Dikstein. 2000. Interaction of TAFII105 with selected p65/RelA dimers is associated with activation of subset of NF-kappaB genes. J. Biol. Chem. 275:18180-18187.

Zeidler, M.P., K. Yokomori, R. Tjian, and M. Mlodzik. 1996. Drosophila TFIIFα-S is up-regulated and required during Mlodzik. Ras-mediated photoreceptor determination. Genes Dev. 10:50–59.