Structural similarity between TAFs and the heterotetrameric core of the histone octamer

Xiaoling Xie, Tetsuro Kokubo, Steven L. Cohen, Urooj A. Mirza, Alexander Hoffmann, Brian T. Chait, Robert G. Roeder, Yoshihiro Nakatani & Stephen K. Burley

Structural similarity between TAFs and the heterotetrameric core of the histone octamer

Xiaoling Xie*†, Tetsuro Kokubo‡#, Steven L. Cohen§, Urooj A. Mirza§, Alexander Hoffmann‖#, Brian T. Chait§, Robert G. Roeder‖, Yoshihiro Nakatani‡ & Stephen K. Burley*††

Laboratories of * Molecular Biophysics, § Mass Spectrometry and Gaseous Ion Chemistry, and † Biochemistry and Molecular Biology, and ‡ Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA
† National Institutes of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

A complex of two TFIID TATA box-binding protein-associated factors (TAFII s) is described at 2.0 Å resolution. The amino-terminal portions of dTAFII42 and dTAFII62 from Drosophila adopt the canonical histone fold, consisting of two short α-helices flanking a long central α-helix. Like histones H3 and H4, dTAFII42 and dTAFII62 form an intimate heterodimer by extensive hydrophobic contacts between the paired molecules. In solution and in the crystalline state, the dTAFII42/dTAFII62 complex exists as a heterotetramer, resembling the (H3/H4)2 heterotetrameric core of the histone octamer, suggesting that TFIID contains a histone octamer-like substructure.

Eukaryotic transcription initiation and its regulation are best understood for genes transcribed by RNA polymerase II (Pol II) with the general transcription factors TFIID, TFIIB, TFIIE, TFIIF, TFIIG/J, TFIH and TFIID (reviewed in refs 1,2). In some cases, this process begins with recognition of the TATA element within the core promoter by the DNA-binding subunit of TFIID (TATA box-binding protein, TBP), forming a multi-protein–DNA complex that coordinates accretion of class II initiation factors and Pol II into a preinitiation complex (PIC) (reviewed in ref. 3). TFIIB is the next general factor to enter the PIC, creating a TFIIB–TFIID–DNA platform that is recognized by Pol II plus TFIIF. In vivo, Pol II transcription depends on TFIIE and TFIH, and possibly TFIIA. Once PIC assembly is complete, and nucleoside triphosphates are present, strand separation occurs to give an open complex, the large subunit of Pol II is phosphorylated, and Pol II initiates transcription and is
released from the promoter. During elongation in vitro, TFIIID can remain bound to the core promoter and support rapid reinitiation of transcription (reviewed in ref. 4). An abbreviated PIC assembly mechanism has also been proposed, following discoveries of various Pol II holoenzymes (reviewed in ref. 5).

The role of TFIIID in eukaryotic transcription has made it the focus of biochemical and genetic studies since its discovery7. DNA binding by human TFIIID was first demonstrated with the adenovirus major late promoter (AdMLP).7 DNase I footprinting studies of the AdMLP and selected human promoters revealed

NATURE • VOL 380 • 28 MARCH 1996
sequence-specific interactions between TFIID and the TATA element that are primarily mediated by TBP (reviewed in ref. 8). In contrast, protection both upstream and downstream of the TATA box is largely sequence independent, displays a nucleosome-like pattern of DNase I hypersensitivity, differs between promoters (for example, AdMLP and the human-shock protein 70 (Hsp70) promoter protect residues -47 to +35 and -35 to -19, respectively), and is induced by some activators (reviewed in ref. 9). TATA box binding by TFIID or TBP precludes packaging of the core promoter with histone proteins (H2A, H2B, H3 and H4). Conversely, core promoter packaging by histone octamers prevents TFIID or TBP binding to the TATA element, effectively repressing transcription (reviewed in ref. 10).

Publication of the sequence of yeast TBP was followed rapidly by the sequences of homologous genes from various eukaryotes and an archaebacterium (amino-acid identities within the carboxy-terminal portion range between 38 and 100%; reviewed in ref. 11), and considerable efforts have been directed towards understanding the mechanisms by which TBP acts in Pol II transcription. Recombinant TBP alone can bind both general and regulatory factors and direct PIC assembly and basal transcription (reviewed in refs 8, 11). Activator-dependent transcription, however, requires TBP and the remaining subunits of TFIID, the TBP-associated factors (TAFs), and some non-TAF coactivators (reviewed in ref. 9). Gene disruption studies of four yeast TAFs demonstrated that they are essential.12,13

Affinity purification of TFIID allowed the identification of a set of phylogenetically conserved TAFs, denoted by their origins and relative molecular mass (reviewed in ref. 9). In both human and Drosophila, TAFs are tightly associated with TBP, providing binding sites for many different transcriptional activators and coactivators that modulate transcription initiation by Pol II by means of specific protein–protein interactions with TFIID (reviewed in ref. 9). Reconstitution of TFIID has been achieved in vitro, providing important insights into the roles of individual subunits as activator-specific targets that facilitate TFIID recruitment.

Primary structure analyses of some TAFs have indicated considerable sequence homology with non-linker histone proteins (Fig. 1)16-19. In Drosophila, dTAF62 (refs 16, 20) and dTAF62 (refs 16, 21) resemble H3 and H4, respectively, and correspond to human hTAF31 (ref. 17) and hTAF80 (refs 17, 21). Both Drosophila and human TAF62 also contain TAFs (dTAF30x and hTAF30x) 16,18,22,23 that are putative histone H2B homologues, but appear to lack histone H2A homologues. The TAF62 nomenclatures of Roeder and Tjian (reviewed in ref. 9) have been adopted for human and Drosophila TFIID, respectively, the only exceptions being dTAF12 and dTAF24 which are denoted dTAF12 and dTAF24 by Tjian. We have previously documented a structural connection between eukaryotic transcription and DNA packaging. The co-crystal structure of the DNA-binding domain of the liver-specific transcription factor hepatocyte nuclear factor (HNF) 3-γ (ref. 24) resembles the structure of the linker histone H5 obtained without DNA25.
Moreover, HNF3-α, a related factor, stabilizes a precisely positioned nucleosomal array in the enhancer of the mouse albumin gene, where it may function as a sequence-specific linker histone.

Here we report the crystal structure of a complex of two RNA polymerase II-specific, TAFs from Drosophila. The 2.0-Å resolution structure reveals that the N-terminal portions of dTAF_{42} and dTAF_{62} are folded into the canonical histone motif. Like histones H3 and H4, dTAF_{42} and dTAF_{62} form an intimate heterodimer. Within the crystal lattice two-fold symmetry generates a dTAF_{42}/dTAF_{62} dimer of heterodimers, which exists in solution and resembles the histone (H3/H4) heterotetramer. We have provided what we believe to be the first high-resolution view of the histone fold, and demonstrated another structural connection between the macromolecular machines responsible for transcription and DNA packaging.

**Crystallization and structure determination**

Following on from the first delineation of histone-like portions of dTAF_{42} and dTAF_{62} from primary sequence alignments, we expressed dTAF_{42}(1–100) and dTAF_{62}(1–91) separately as inclusion bodies in Escherichia coli, with the latter as a glutathione S-transferase (GST) fusion protein. Equimolar amounts of denatured dTAF_{42}(1–100) and dTAF_{62}(1–91) were co-renatured and purified to homogeneity. The resulting dTAF_{42}(1–100)/dTAF_{62}(1–91) complex preparation was monodisperse and tetrameric, as judged by dynamic light scattering (data not shown), and largely α-helical, as judged by circular dichroism spectroscopy (data not shown).

Initial crystallization trials immediately yielded small crystals, which diffracted weakly to 4–5 Å and resisted attempts to improve
their properties. This result underscores the fact that, although monodisperse preparations of globular proteins and protein-DNA complex are excellent candidates for crystallization trials, monodispersity does not guarantee diffraction-quality crystals. Reasoning that flexible portions of one or both of these truncated proteins might be interfering with formation of a well-ordered crystal lattice, we searched for a proteolytic limit digest of the dTAF₂₄(1–100)/dTAF₆₂(1–91) complex. (For the results of a retrospective analysis of our crystallization trials with the Max-DNA complex, see ref. 29). Proteolysis was combined with mass spectrometry to rapidly obtain accurate cleavage maps (Fig. 1a). These data suggested that three of the four termini of dTAF₂₄(1–100) and dTAF₆₂(1–91) had been chosen incorrectly. New E. coli expression vectors encoding dTAF₂₄(11–95) and GST-dTAF₆₂(1–82) were assembled, allowing the production of another truncated form of the dTAF₂₄/dTAF₆₂ complex. Crystallization trials immediately yielded large crystals in the tetragonal space group P4₁2₁2 (a = 59.8 Å, c = 111.0 Å), with one heterodimer in the asymmetric unit that diffracted to at least 1.4 Å resolution (Fig. 1). Trypsin treatment of (H3/H4)₅ yielded a stable complex with improved crystallization properties.

The dTAF₂₄/dTAF₆₂ X-ray structure was determined at 2.4 Å resolution by a combination of multiwavelength anomalous dispersion (MAD) and multiple isomorphous replacement (MIR) (Fig. 1, Table 1), followed by X-PLOY refinement to 2.0 Å resolution (Table 1). Our model consists of residues 17–86 of dTAF₂₄, residues 1–70 of dTAF₆₂, 125 water molecules, and 7 Zn²⁺ ions (Fig. 2), giving an R-factor of 19.8% with Rfree = 24.4%. The electron density for the polypeptide backbone is everywhere continuous at 1.2σ in a (2|Fₒ| – |Fₑ|) difference Fourier synthesis (Fig. 3). PROCHECK revealed 1/140 unfavourable (Φ, ψ) combinations, and main-chain and side-chain structural parameters consistently better than those expected at 2.0-Å resolution (overall G-factor = 0.5). A full account of the refinement at the diffraction limit will be published elsewhere.

**Structures of dTAF₂₄ and dTAF₆₂**

The X-ray structures of the histone-like regions of dTAF₂₄ and dTAF₆₂ are illustrated in Fig. 2. (For clarity, the TAFII fragments used in this study are denoted dTAF₂₄(11–95) and dTAF₆₂(1–82). The portions of these two polypeptide chains visualized in the final electron-density map are denoted dTAF₂₄(17–86) and dTAF₆₂(1–70), with their histone-like regions denoted dTAF₂₄(22–83) and dTAF₆₂(9–70), respectively. As predicted, the N-terminal portions of both dTAF₂₄ and dTAF₆₂ adopt the canonical histone motif, consisting of a long central α-helix flanked on each side by a short α-helix (Fig. 2). The truncated form of dTAF₂₄ used for crystallization does not include the portion of the polypeptide chain corresponding to the additional H3 N-terminal α-helix, observed in the histone octamer structure. Throughout this paper, we describe structural details using α1, L1, α2, L2 and α3 to denote the segments within the histone fold (italics/sloping Greek denote dTAF₆₂). Comparisons of the histone-homology regions of the TAFII structures with histones H3 and H4 revealed the following root-mean-square deviations (r.m.s.d.s) between α-carbon atomic positions: dTAF₂₄(22–83) versus dTAF₆₂(9–70), 1.6 Å; dTAF₂₄(22–83) versus H3(68–130), 1.6 Å; and dTAF₆₂(9–70) versus H4(31–92), 1.6 Å, respectively. These values compare favourably with those obtained by comparing H2A, H2B, H3 and H4 with one another (see Table 1 in ref. 33). Presumably these relatively large r.m.s.d.s reflect differences in the trajectory of α2. In H3 the α2 helix is linear, but is bent near its C terminus in dTAF₂₄; the converse is true for H4 and dTAF₆₂.

The canonical histone fold has been postulated to arise from gene duplication, giving two helix-strand–helix (HSH) segments (HSH1 and HSH2 refer to the N- and C-terminal halves of the histone fold, respectively). The results of a detailed comparison of HSH1 from dTAF₂₄(22–51), dTAF₆₂(9–38), H3(68–98) and H4(31–60), and HSH2 from dTAF₂₄(52–83), dTAF₆₂(39–70), H3(99–130) and H4(61–92), are shown in Fig. 1. The HSH2 fold is very similar, displaying r.m.s.d.s between α-carbon atomic positions of 0.7–1.0 Å. In contrast, the HSH1 fold is less similar, with the corresponding r.m.s.d.s ranging from 1.4 to 1.9 Å. This marked difference between HSH1 and HSH2 may reflect the critical role of HSH2 in mediating interactions.

**FIG. 4 Electrode-spray–ionization mass spectrum of the co-renatured dTAF₂₄(11–95)/dTAF₆₂(1–82) complex dissolved in water, showing peaks corresponding to multiply charged states of the non-covalent dTAF₂₄(11–95)/dTAF₆₂(1–82) heterodimer (D) plus the dTAF monomers (M1 and M2).**

**FIG. 3 a, Stereo drawing of the heterodimer interface, showing the interactions between dTAF₂₄(17–86) and dTAF₆₂(1–70).** Critical hydrophobic residues at the interface are labelled with single-letter amino-acid code. The water molecule bridging Tyr 52 to Asp 35 and Thr 81 to Asp 35 can be seen above the OH group of Tyr 52. b, Stereo drawing of the 2.0 Å resolution (2|Fₒ| – |Fₑ|) difference Fourier synthesis for the portion of the TAF₂₄ complex depicted in a.
between heterodimers within the histone octamer crystal structure or between TAF₆₂ heterodimers and homodimers.

Differential scanning calorimetry demonstrated a single cooperative unfolding transition for the H3/H4 complex. By analogy, the histone-like portions of dTAF₆₂ and dTAF₆₄ are probably unable to assume their full native folds in the absence of one another, because they too demonstrate only a small number of long-range intramolecular contacts. Residues involved in these contacts have been referred to as 'self' residues (denoted with 's' in Fig. 1a). For dTAF₆₂(17–86) these intramolecular interactions include x₁–x₂, x₁–L₁, x₂–L₂, x₂–x₃ and L₂–x₃; for dTAF₆₄(17–70), they include x₁–x₂, x₁–L₁, x₂–L₂, x₂–x₃ and L₂–x₂ (italics/sloping Greek are used throughout this paper to refer to dTAF₆₄).

**Structure of dTAF₆₂/dTAF₆₄**

The structure of the dTAF₆₂(17–86)/dTAF₆₄(1–70) heterodimer is shown in Fig. 2. As in the H3/H4 heterodimer, also illustrated in Fig. 2, the two histone folds interact with one another in a head-to-tail fashion. Stabilizing contacts between dTAF₆₂(17–86) and dTAF₆₄(1–70) are largely hydrophobic, spanning the entire length of both molecules, and are conserved with H3 and H4 (Fig. 1). Binary complex formation buries about 3,390 Å² of solvent-accessible surface area (56% of the buried surface is hydrophobic, with the remainder either polar or charged). The residues involved in heterodimer contacts have been referred to as 'pair' residues (denoted with 'p' in Fig. 1a). dTAF₆₂(17–86)/dTAF₆₄(1–70) heterodimer interactions involve each segment of the two polypeptide chains, including x₁–x₂, x₁–L₂, L₁–x₂, L₁–L₂, L₁–x₃, x₂–x₃, x₂–x₂, x₂–L₁, x₂–L₂, x₂–x₂, x₂–L₁, x₂–L₂. Asymmetry in these interactions probably explains why H3, H4, dTAF₆₂ and dTAF₆₄ do not show any significant propensity to form homodimers. Electro spray ionization mass spectrometry was also used to examine the stability of the dTAF₆₂(11–95)/dTAF₆₄(1–82) heterodimer. Under moderately gentle electrospray conditions, a pure water solution of co-renatured dTAF₆₂(11–95)/dTAF₆₄(1–82) exhibits strong multiply charged peaks corresponding to the non-covalent heterodimer and the individual monomers (Fig. 4).

**Structure of (dTAF₆₂/dTAF₆₄)**

The structure of the (dTAF₆₂(17–86)/dTAF₆₄(1–70)) heterotetramer is shown in Fig. 5. As in the histone octamer, the symmetry axis within the TAF₆₂ tetramer coincides with a crystallographic two-fold. Interactions involving the HSH2 portion of the H3 homologue, dTAF₆₂(17–86), stabilize the heterotetramer, burying about 670 Å² of solvent-accessible surface area (48% of the buried surface is hydrophobic, with the remainder either polar or charged). These values are typical for protein-protein complexes (reviewed in ref. 36), and are consistent with the equilibrium dissociation constant of 10⁻⁸ M (Fig. 1). Residues involved in dTAF₆₂(17–86)/dTAF₆₄(17–86) interactions in the crystalline state are denoted with an asterisk in Fig. 1a, and correspond to conserved residues in H3.

We presume that the same residues (Asp 59, Val 62, Tyr 63, His 66, Leu 79, Glu 82 and Val 83) stabilize the solution form of the (dTAF₆₂/dTAF₆₄) heterotetramer, but cannot be certain without experimental confirmation. The proteolysis–mass spectrometry results provide indirect support for this prediction. Enzymatic cleavage at three sites (Tyr 63, Arg 68 and Asp 76) within the HSH2 region of dTAF₆₂(1–100) is slow compared with cleavage events outside the histone-homology region (Fig. 1a). Two of these three sites are solvent inaccessible in the crystallographic heterotetramer, and all three are completely exposed in the crystallographic heterodimer.

**TFIID assembly**

Our biophysical study of the dTAF₆₂/dTAF₆₄ complex suggests that TFIID contains a (dTAF₆₂/dTAF₆₄), heterotetramer, which could interact with H2B-like TAF₁₆ to form a histone-like octamer. Compelling, albeit indirect, support for these predictions comes from the results of recent studies of the human TAF₆₂ homologue of histone H2B (hTAF₆₂)⁹. The histone-like region of hTAF₆₂ alone allows assembly of TFIID in vivo, the measured hTAF₆₂:TBP ratio in TFIID is 4:1, and an octamer-like pattern of protein–protein interactions has been documented for hTAF₆₂(31) (homologous to H3 and dTAF₆₂), hTAF₆₂(80) (homologous to H4 and dTAF₆₄), and hTAF₆₂(20) (homologous to H2B and dTAF₆₄(30))¹⁰. In addition, the histone-like regions of dTAF₆₂ and dTAF₆₄ contain HSH2 residues homologous, respectively, to the H4 and H2B residues responsible for mediating H4–H2B interactions in the histone octamer structure¹¹ (Fig. 1a). Protein–protein interactions have also been demonstrated between hTAF₆₂ and TBP, and between hTAF₆₂ and TBP, hTAF₆₁₃ and hTAF₆₅ (ref. 18). Thus TFIID may contain a histone-like octameric TAF₆₂ substructure (hTAF₆₂(20)−(hTAF₆₂(31)−dTAF₆₄(80))−(hTAF₆₄(20)−(dTAF₆₂(20)−(dTAF₆₂(30))−(dTAF₆₄(30))))−(dTAF₆₄(30))) anchored to TBP, hTAF₆₁₃ (or dTAF₆₁₁₀) and hTAF₆₅ (there is no known Drosophila homologue for this human TAF₆₂);¹² There are other TBP–TAF₆₂ and TAF₆₂–TAF₆₄ interactions (reviewed in ref. 9) that probably stabilize the remainder of the TFIID complex.

The co-crystal structure of core TFIIB recognizing the preformed TBP–DNA complex showed the second step of PIC assembly, explaining stabilization of the TBP–TATA element
complex by TFIIIB. With TFIIIB instead of TBP, however, the situation may be more complicated. Affinity chromatography demonstrated that dTARF42 and its homologue hTARF31 interact with TFIIH. Similar methods identified interactions between hTARF80 (the human homologue of dTARF62) and two other general transcription factors TFIIF and TFIIIE. Although the functional significance of these interactions has not been established, it seems likely that they increase PIC stability, thereby facilitating recruitment of the corresponding general transcription factors. They may also do so by means of conformational changes induced in the TFIIH complex by interactions with transcriptional activators or non-TAFi coactivators.

Our current idea of activator–TFIID interactions suggests that the TAFiSs constitute a large multiprotein complex that sits on top of TBP, and integrates signals from many activators and non-TAFi coactivators. Histone-like TAFiF/activator contacts include: interactions between the isoleucine-rich activation domain of NTF-1 and dTARF62 (ref. 14); interactions of the p53 and NFκB/p65 activation domains with dTARF42 (or its homologue hTARF31) and dTARF62 (or its homologue hTARF80)30,38 (M. Guermah & R. G. Roeder, unpublished observations); and the interactions of the VP16 activation domain with dTARF42 (or its homologue hTARF31)31,38. We do not know if any of the interactions between TAFiSs and histone homologues and various transcriptional activators map to the histone-like regions. The activators may target portions of the polypeptide chains of dTARF42 and dTARF62 that were removed to facilitate crystallization.

Another significant aspect of activator–TAFiS interactions concerns induced conformational changes in TFIID and TFIID-promoter complexes. Activator-induced changes have been demonstrated in TFIIH-promoter complex46, manifested by downstream extension of the TFIIH footprint well beyond the transcription start site that was correlated with increased recruitment of other general factors. Later studies with highly purified TFIIH have confirmed qualitative and quantitative effects of activators on TFIIH binding, sometimes requiring TFIIA44. Photoaffinity DNA–protein crosslinking studies have also revealed large changes in TAFiS–DNA contacts in response to interactions of TFIIA with promoter-bound TFIID (T. Oegelshlager & R. G. Roeder, unpublished observations). Together, these data suggest that binding of transcriptional activators or coactivators can cause substantial rearrangements in the relative disposition of TAFiS subunits and DNA. In this context, it might be imagined that activator-induced changes in TAFiS–DNA interactions allow the presumptive histone octamer–like substructure within TFIIH to engage DNA. We speculate that such a structure might be involved in stabilizing an activator–TFIID–promoter complex, yielding a stereospecific nucleoprotein assembly that can support transcriptional activation (reviewed in ref. 45). We do not, however, predict that this putative octamer substructure would necessarily be encircled by a piece of DNA 146 base pairs long, as in the nucleosome core particle (reviewed in refs 46–49). It is also possible that TFIIH merely exploits the histone fold to mediate protein–protein contacts between various TAFiSs.

**Conclusion**

We believe that the 2.0–Å resolution structure of the dTARF42/dTARF62 complex provides the first structural knowledge of RNA polymerase II TAFiSs, and the first high-resolution structure of the histone fold. Our work also provides a starting point for further crystallographic, biochemical and genetic studies of TFIIH assembly and its role in Pol II-mediated transcription activation. Finally, studies of the three known histone-like TAFiS suggest an evolutionary relationship between the histone octamer and a TFIIH substructure that may be functionally significant.