Cell Research (2014) 24:1433-1444.

www.nature.com/cr

ORIGINAL ARTICLE

Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module

Hui Wang[1, 2], Elizabeth C Curran[1], Thomas R Hinds[1, 2], Edith H Wang[1], Ning Zheng[1, 2]

[1]Department of Pharmacology, [2]Howard Hughes Medical Institute, Box 357280, University of Washington, Seattle, WA 98195, USA

The general transcription factor IID (TFIID) initiates RNA polymerase II-mediated eukaryotic transcription by nucleating pre-initiation complex formation at the core promoter of protein-encoding genes. TAF1, the largest integral subunit of TFIID, contains an evolutionarily conserved yet poorly characterized central core domain, whose specific mutation disrupts cell proliferation in the temperature-sensitive mutant hamster cell line *ts*13. Although the impaired TAF1 function in the *ts*13 mutant has been associated with defective transcriptional regulation of cell cycle genes, the mechanism by which TAF1 mediates transcription as part of TFIID remains unclear. Here, we present the crystal structure of the human TAF1 central core domain in complex with another conserved TFIID subunit, TAF7, which biochemically solubilizes TAF1. The TAF1-TAF7 complex displays an inter-digitated compact architecture, featuring an unexpected TAF1 winged helix (WH) domain mounted on top of a heterodimeric triple barrel. The single TAF1 residue altered in the *ts*13 mutant is buried at the junction of these two structural domains. We show that the TAF1 WH domain has intrinsic DNA-binding activity, which depends on characteristic residues that are commonly used by WH fold proteins for interacting with DNA. Importantly, mutations of these residues not only compromise DNA binding by TAF1, but also abrogate its ability to rescue the *ts*13 mutant phenotype. Together, our results resolve the structural organization of the TAF1-TAF7 module in TFIID and unveil a critical promoter-binding function of TAF1 in transcription regulation.

***Keywords*:** Transcription; TFIID; TAF; winged helix; DNA binding; cell cycle progression

Cell Research (2014) **24**:1433-1444. doi:10.1038/cr.2014.148; published online 21 November 2014

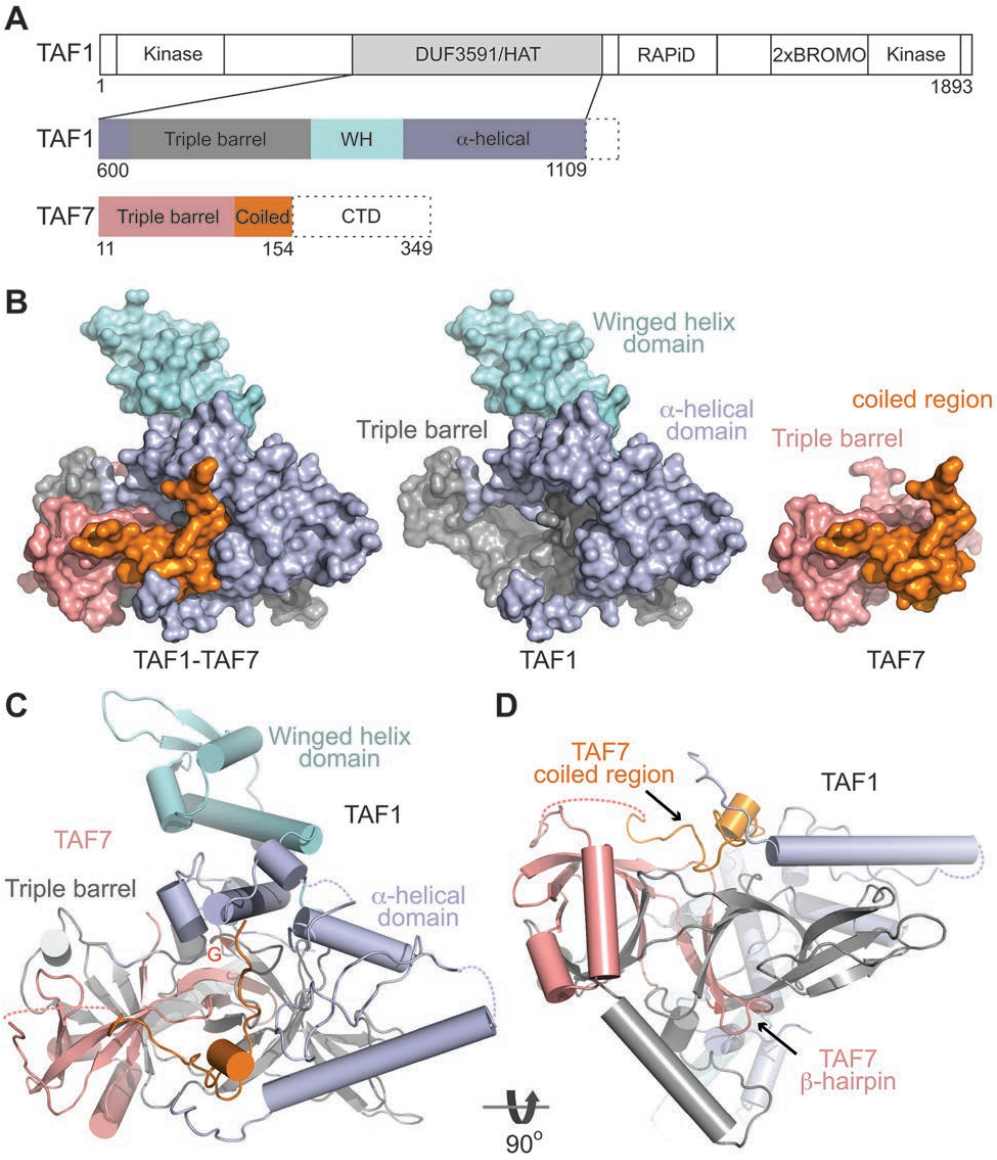
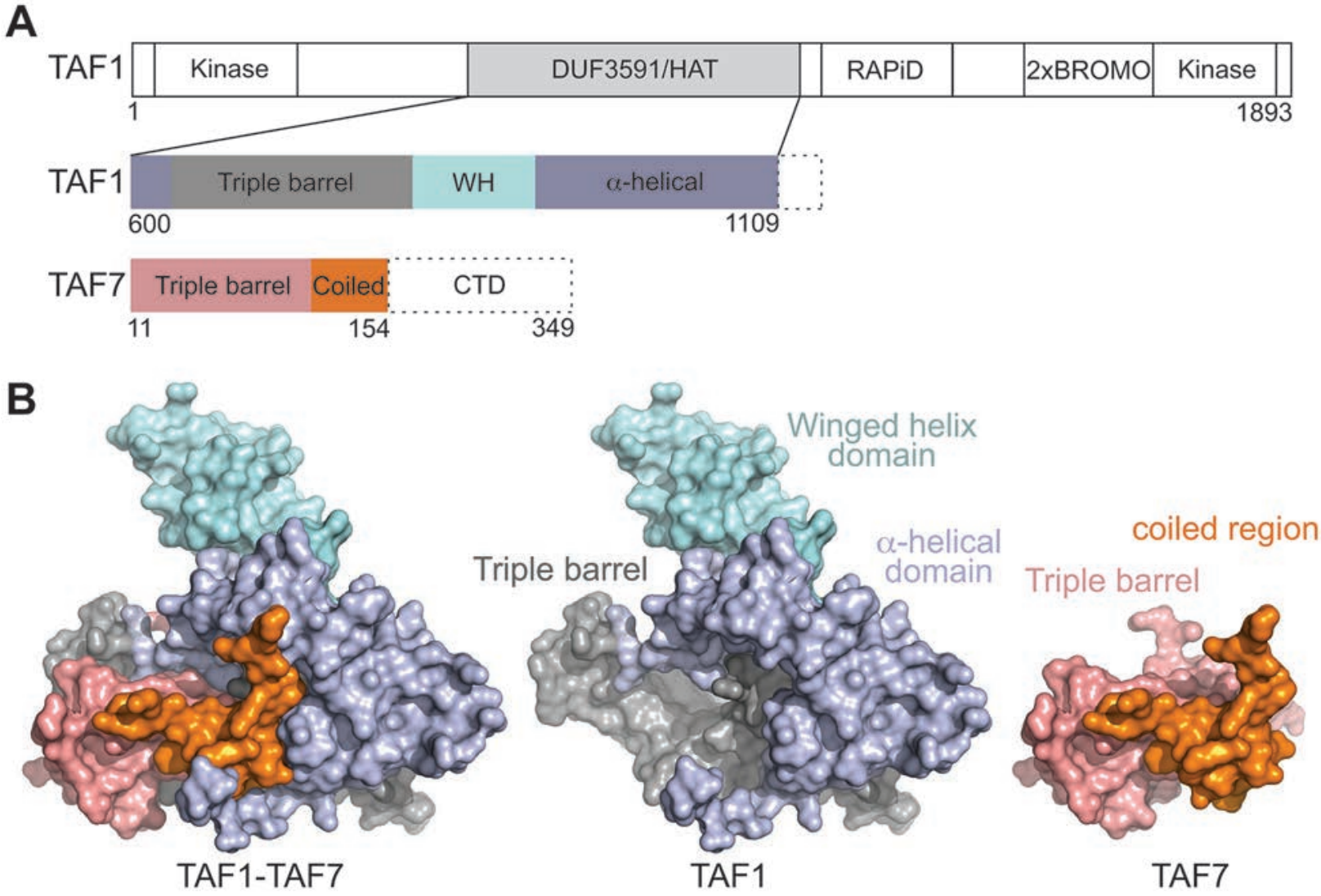


Figure 1 Overall structure of the TAF1-TAF7 complex. **(A)** Domain organization of TAF1 and the TAF1 and TAF7 fragments crystallized by the *in situ* proteolysis method. Different structural domains of the crystallized TAF1 and TAF7 fragments are labeled and colored. Regions outlined by dotted lines represent the proteolytically removed segments. **(B)** Surface representations of the TAF1-TAF7 complex with structural domains colored with the same scheme as shown in **A**. TAF1 and TAF7 are separately shown on the right from the same angle. **(C**, **D)** Ribbon diagram of the TAF1-TAF7 complex structure in two orthogonal views. The TAF1 and TAF7 protein domains are colored with the same scheme as in **A**. The TAF1 G716 residue mutated in the *ts*13 hamster cell line is colored in red and indicated by a letter "G". Dashed lines represent loop regions that are not visible in the crystal structure.


A
TAF1
Kinase
DUF3591/HAT
RAPiD
2xBROMO
Kinase
1
1893
TAF1
Triple barrel
WH
α-helical
600
1109
TAF7
Triple barrel
Coiled
CTD
11
154
349
B
Winged helix domain
α-helical domain
Triple barrel
coiled region
Triple barrel
TAF1-TAF7
TAF1
TAF7

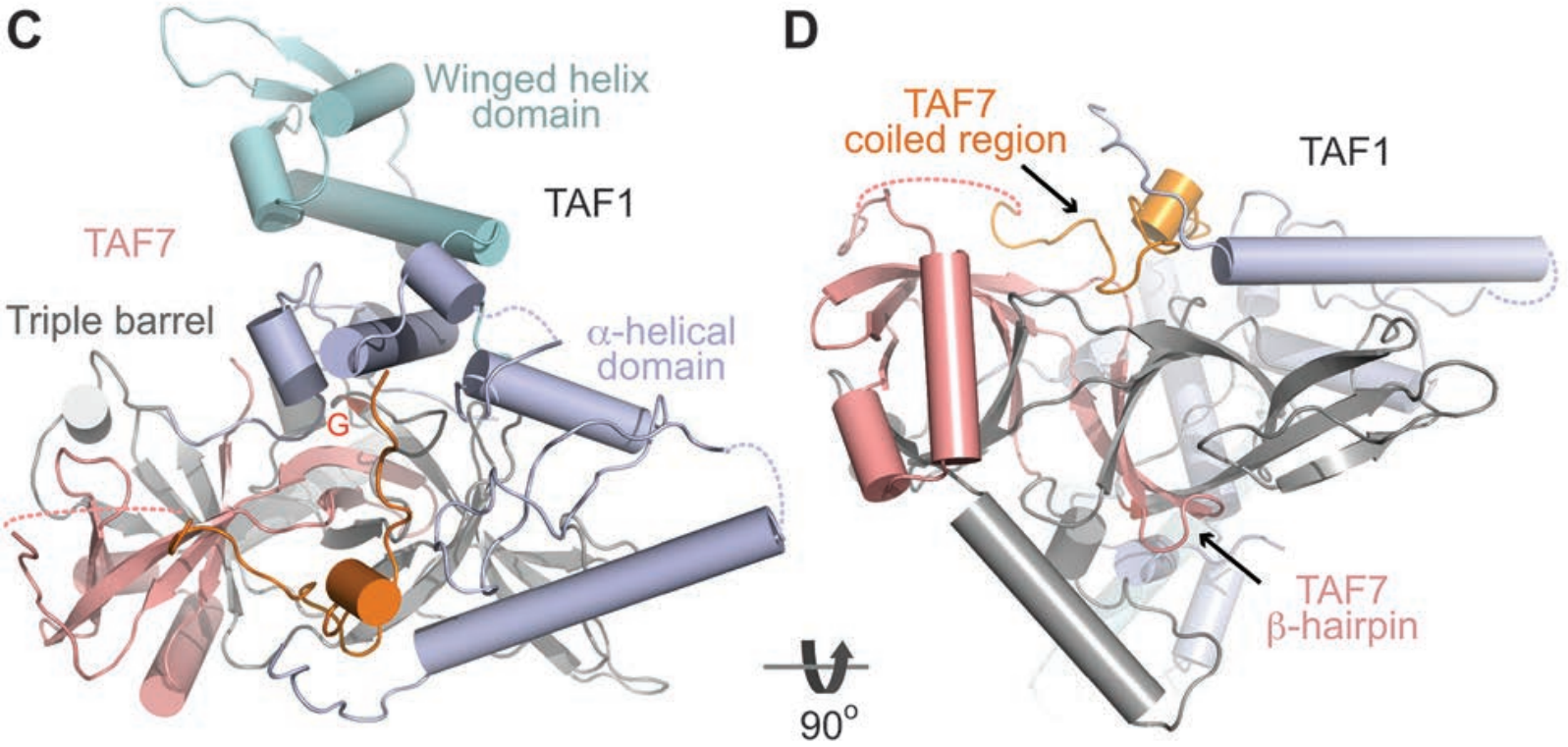


Figure 1 Overall structure of the TAF1-TAF7 complex. **(A)** Domain organization of TAF1 and the TAF1 and TAF7 fragments crystallized by the *in situ* proteolysis method. Different structural domains of the crystallized TAF1 and TAF7 fragments are labeled and colored. Regions outlined by dotted lines represent the proteolytically removed segments. **(B)** Surface representations of the TAF1-TAF7 complex with structural domains colored with the same scheme as shown in **A**. TAF1 and TAF7 are separately shown on the right from the same angle. **(C**, **D)** Ribbon diagram of the TAF1-TAF7 complex structure in two orthogonal views. The TAF1 and TAF7 protein domains are colored with the same scheme as in **A**. The TAF1 G716 residue mutated in the *ts*13 hamster cell line is colored in red and indicated by a letter "G". Dashed lines represent loop regions that are not visible in the crystal structure.

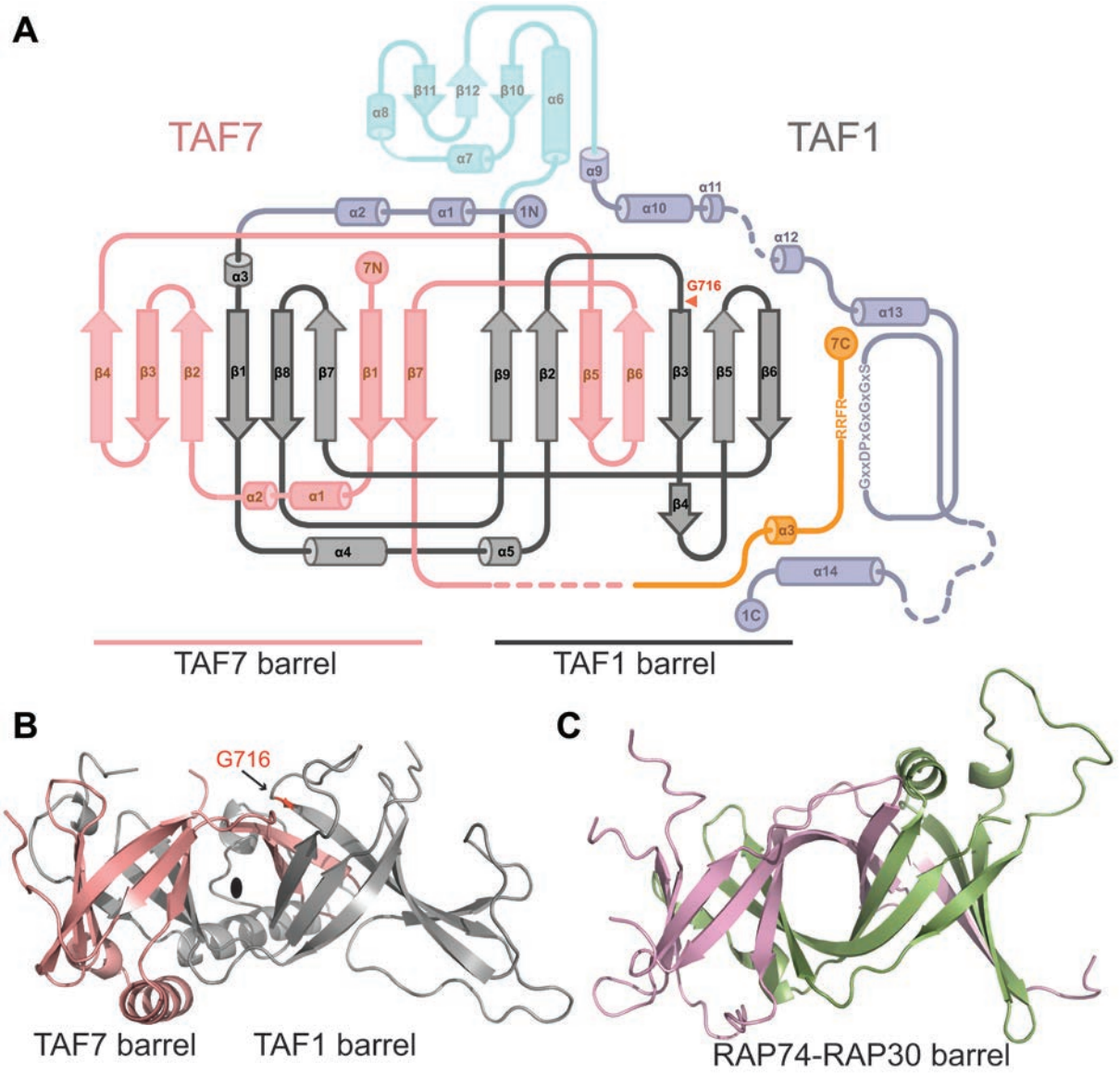


Figure 2 The TAF1-TAF7 heterodimeric triple barrel. **(A)** Topology diagram of the TAF1-TAF7 triple barrel in the context of the entire crystallized complex of TAF1 and TAF7. Dashed lines represent loop regions not visible in the crystal structure. The Gly-rich motif of TAF1 and the Arg-rich motif of TAF7 are highlighted. **(B)** Ribbon diagram of the TAF1-TAF7 triple barrel core. The pseudo two-fold symmetry axis relating the two distal barrels is indicated by the black oval. The residue G716, which is mutated in the *ts*13 hamster cell line, is colored in red and indicated. **(C)** Ribbon diagram of the RAP74-RAP30 triple barrel.

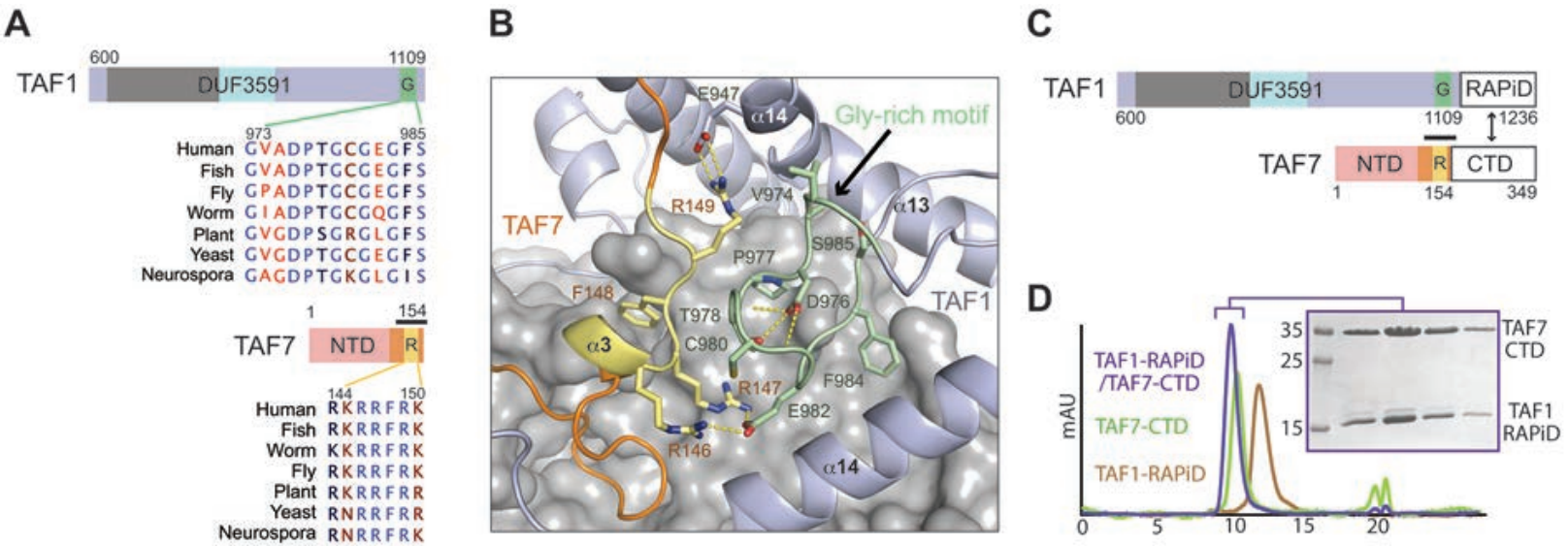


Figure 3 TAF1-TAF7 interactions between conserved loop motifs and C-terminal regions. **(A)** Schematic diagram of the crystallized TAF1 DUF3591 domain and TAF7 NTD with their conserved Gly-rich and Arg-rich sequence motif, respectively. The interacting TAF1 Gly-rich motif and TAF7 Arg-rich motif are colored pale green and bright yellow, and labeled “G” and “R”, respectively. Black bar above TAF7 represents the 19-amino acid sequence, which has been documented as a critical TAF1-binding region [18]. **(B)** A close-up view of the interactions between the conserved TAF1 Gly-rich motif and TAF7 Arg-rich motif, which are colored in the same scheme as shown in **A**. Dashed lines indicate intermolecular hydrogen bonds and salt bridges. The triple barrel in the complex is shown by surface representation. **(C)** Schematic diagram of the interactions between the TAF1 RAPiD domain and the TAF7 CTD. **(D)** Size exclusion chromatography elution profiles of the isolated TAF1-RAPiD domain (brown), the TAF7 CTD (green), and a mixture of the two (purple). Similar to the full-length protein, the TAF7 CTD migrates as a larger species on SDS-PAGE.

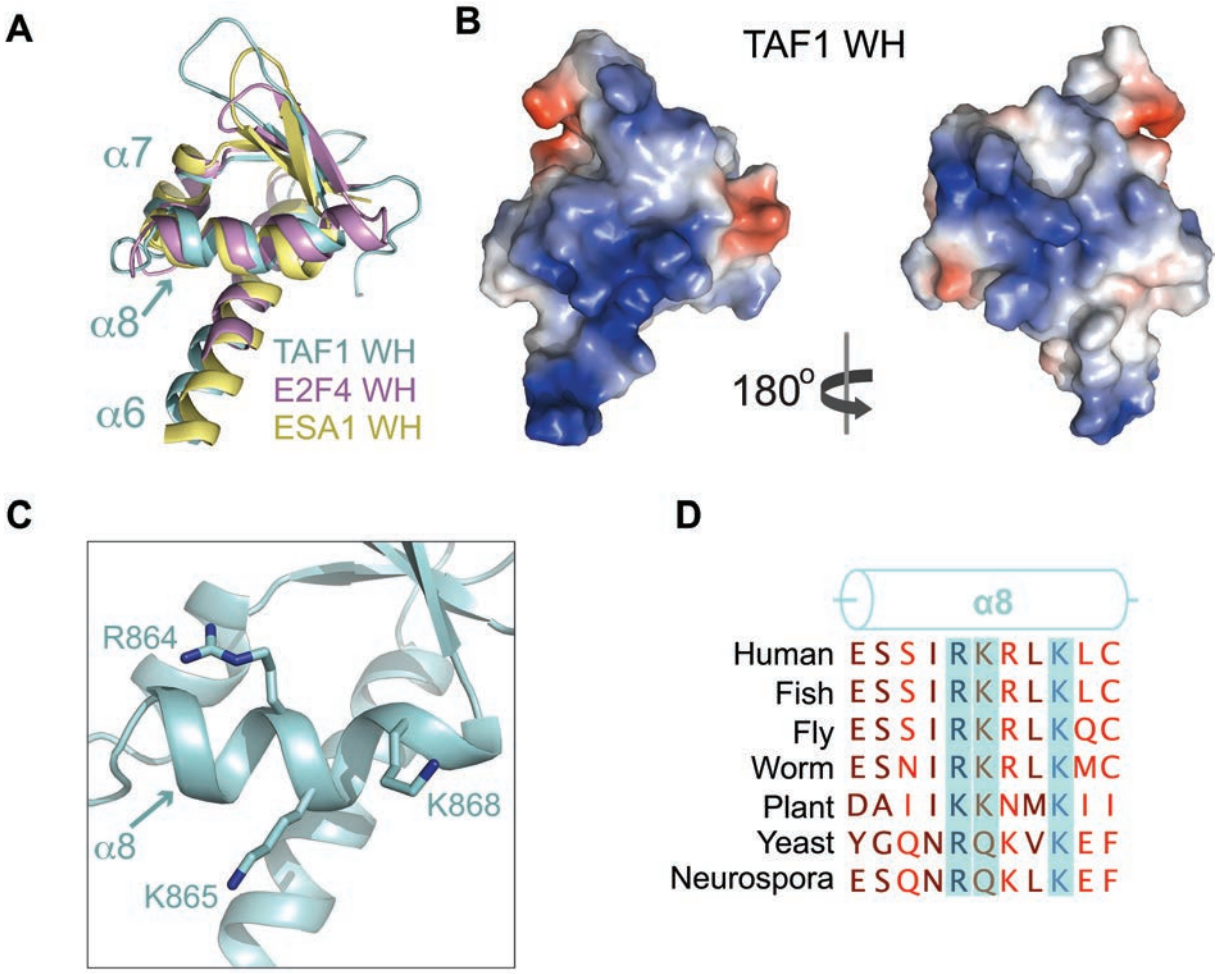


Figure 4 The winged helix (WH) domain of TAF1. **(A)** Superposition of the WH domains of TAF1, the transcription factor E2F4, and the yeast HAT ESA1 protein. **(B)** Electrostatic potential surface of the TAF1 WH domain viewed from the same angle as in **A** and 180° away. The surface colors are clamped between red (–83.5kTe^{-1}) and blue (+83.5kTe^{-1}). **(C)** A close-up view of the α8 helix of TAF1 with three solvent-exposed basic residues. **(D)** Sequence alignment of α8 helix in the TAF1 WH domain. The three basic residues mutated in our studies are highlighted in cyan.

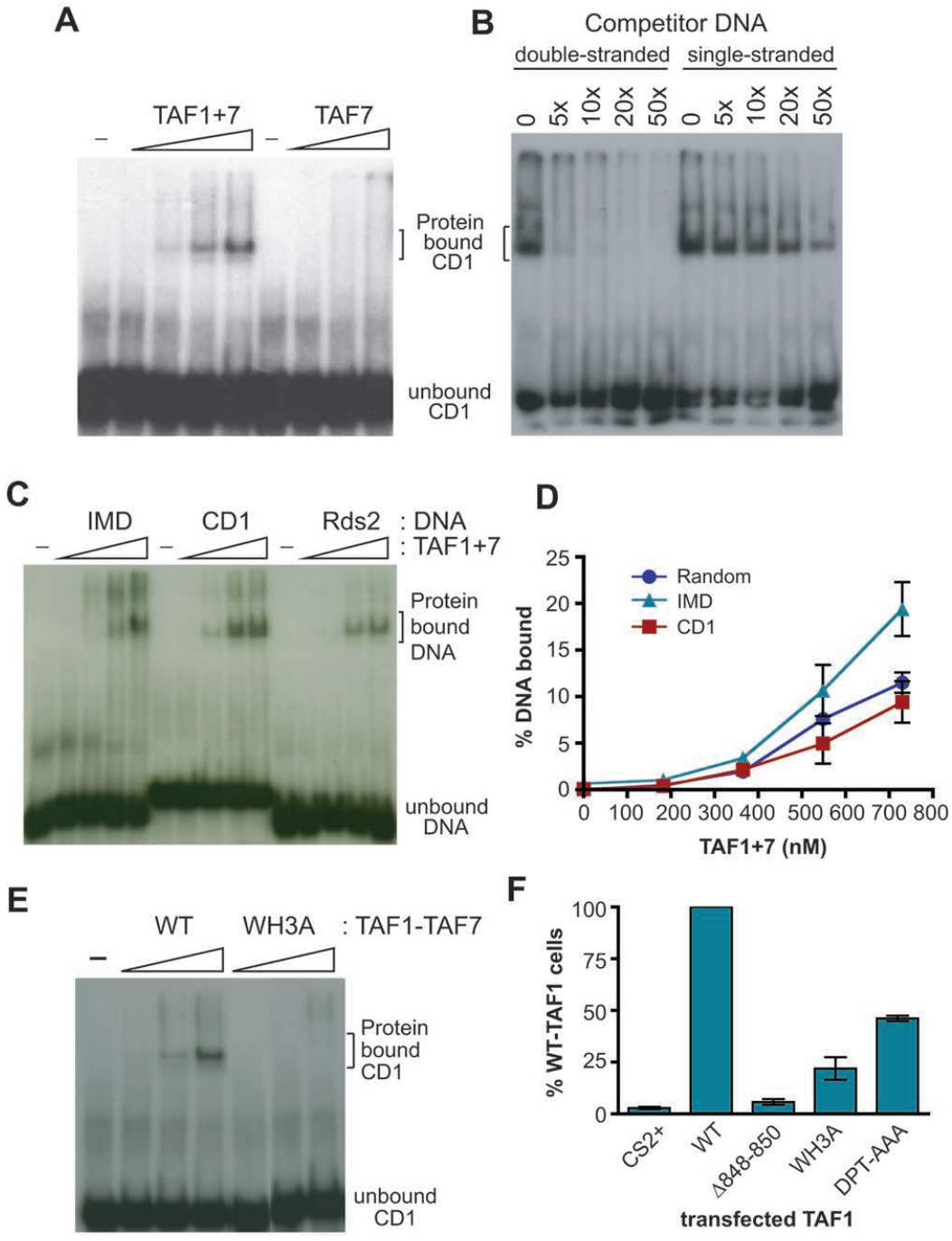
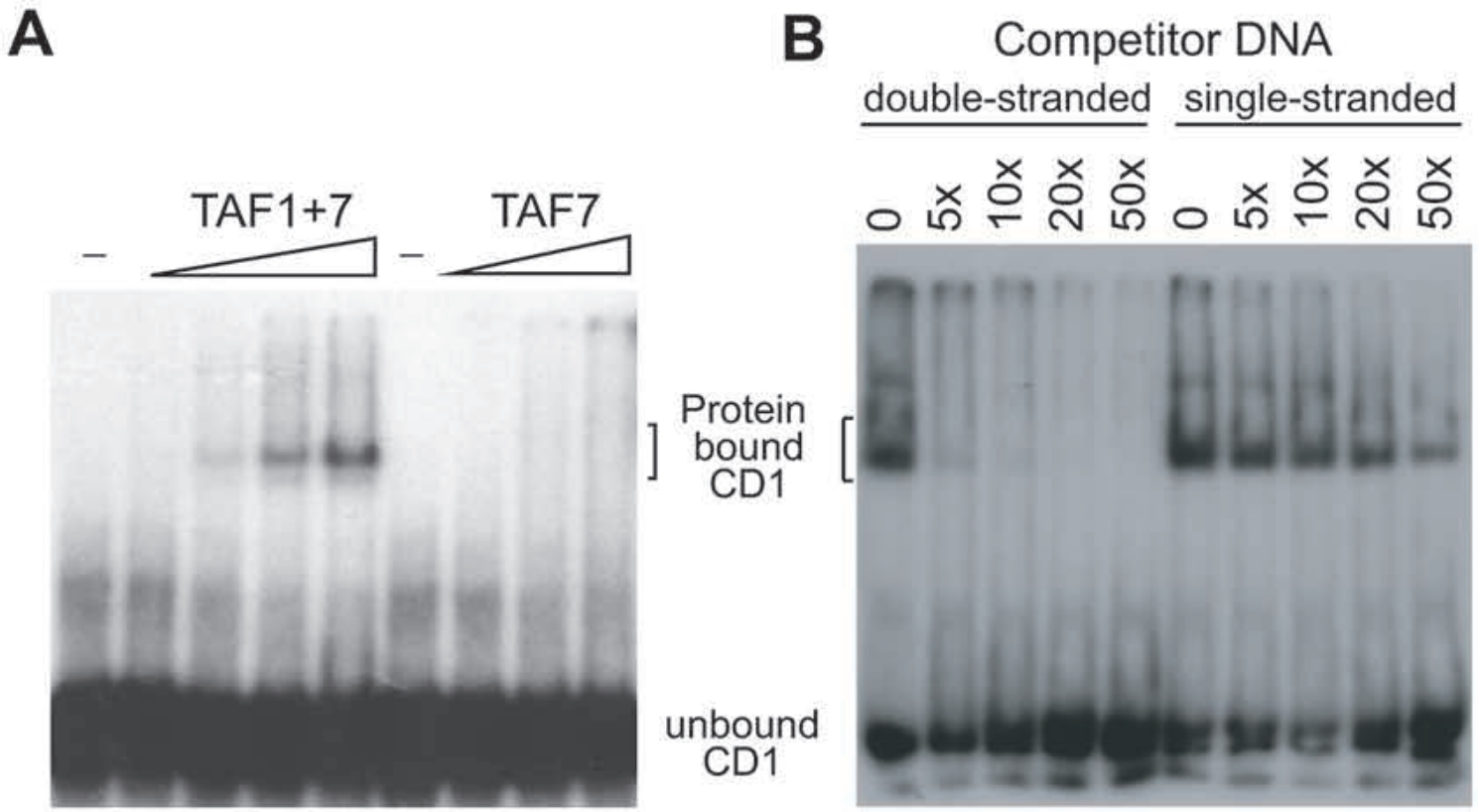

A
TAF1+7
TAF7
–
Protein bound CD1
unbound CD1
B
Competitor DNA
double-stranded
single-stranded
0
5x
10x
20x
50x
C
IMD
CD1
Rds2
: DNA
: TAF1+7
Protein bound DNA
unbound DNA
D
% DNA bound
25
20
15
10
5
0
Random
IMD
CD1
0 100 200 300 400 500 600 700 800
TAF1+7 (nM)
E
WT
WH3A
: TAF1-TAF7
Protein bound CD1
unbound CD1
F
% WT-TAF1 cells
100
75
50
25
0
CS2+
WT
Δ848-850
WH3A
DPT-AAA
transfected TAF1

Figure 5 DNA binding and *ts*13 cell complementation activities of WT and mutant TAF1 proteins. **(A)** TAF1 displays DNA-binding activity. EMSA was carried out with increasing concentrations of the indicated purified proteins. Binding reactions were resolved on native 5% polyacrylamide and subjected to autoradiography. The positions of protein-bound and -unbound ^{32}P-labeled cyclin D1 (CD1) core promoter fragments are shown. **(B)** TAF1 binds to double-stranded DNA. TAF1-TAF7 (22.5 pmoles) was incubated with radiolabeled CD1 probes in the presence of the indicated molar excess amount of unlabeled double-stranded or single-stranded CD1. Reaction products were analyzed and detected as described in **A**. **(C)** Minimal sequence-specificity for TAF1 DNA binding. Increasing concentrations of TAF1-TAF7 complex was subjected to EMSA using different radiolabeled DNA probes. IMD: initiator and downstream region of super core promoter, CD1: cyclin D1 core promoter, Rds2: random double-stranded DNA sequence. Migrations of protein-bound and -unbound DNA fragments are indicated. **(D)** Quantitation of DNA binding to the DNA probes described in **C**. Regions containing the bound and unbound radiolabeled DNA probes were excised from the dried gel and quantified by liquid scintillation. The percentage of total counts that the shifted complex represented was calculated for each sample. The graphed results are the average from several independent experiments, IMD: $n = 3$; CD1: $n = 4$; Rds2: $n = 3$. Error bars represent SEM. **(E)** WH domain mutations abolish DNA-binding activity of TAF1. TAF1-TAF7 complexes assembled in insect cells with WT-TAF1 or TAF1-WH mutant (WH3A) were analyzed by EMSA. Protein-bound and -unbound ^{32}P-labeled CD1 fragments were detected by autoradiography. **(F)** TAF1 mutations compromise growth complementation efficiency in *ts*13 cells. Mutant *ts*13 cells were transfected with the indicated CS2+ control or TAF1 expression plasmid at 33.5 °C and subsequently shifted to 39.5 °C. After 48-72 h at the elevated temperature, the number of viable cells was counted and expressed as a percent relative to cells transfected with WT-TAF1, given a value of 100%. The bar graph depicts the average from 4 independent complementation experiments. Error bars, SEM. CS2+, empty vector; WT, WT-TAF1; Δ848-850, deletion of residues 848-850; WH3A: R864/K865/K868 all changed to alanine; DPT-AAA: D976/P977/T978 all changed to alanine.


A
TAF1+7
TAF7
–
–
Protein bound CD1
unbound CD1
B
Competitor DNA
double-stranded
single-stranded
0
5x
10x
20x
50x
0
5x
10x
20x
50x

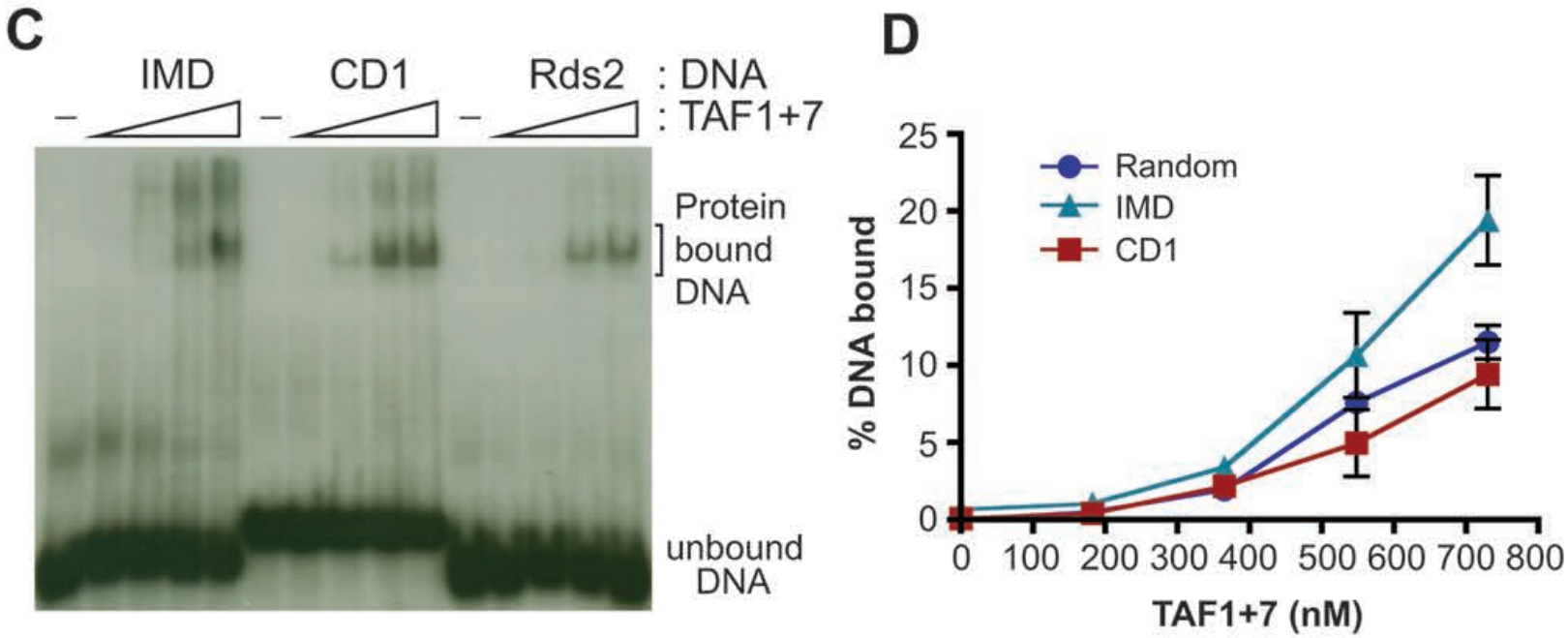


Figure 5 DNA binding and *ts*13 cell complementation activities of WT and mutant TAF1 proteins. **(A)** TAF1 displays DNA-binding activity. EMSA was carried out with increasing concentrations of the indicated purified proteins. Binding reactions were resolved on native 5% polyacrylamide and subjected to autoradiography. The positions of protein-bound and -unbound ^{32}P-labeled cyclin D1 (CD1) core promoter fragments are shown. **(B)** TAF1 binds to double-stranded DNA. TAF1-TAF7 (22.5 pmoles) was incubated with radiolabeled CD1 probes in the presence of the indicated molar excess amount of unlabeled double-stranded or single-stranded CD1. Reaction products were analyzed and detected as described in **A**. **(C)** Minimal sequence-specificity for TAF1 DNA binding. Increasing concentrations of TAF1-TAF7 complex was subjected to EMSA using different radiolabeled DNA probes. IMD: initiator and downstream region of super core promoter, CD1: cyclin D1 core promoter, Rds2: random double-stranded DNA sequence. Migrations of protein-bound and -unbound DNA fragments are indicated. **(D)** Quantitation of DNA binding to the DNA probes described in **C**. Regions containing the bound and unbound radiolabeled DNA probes were excised from the dried gel and quantified by liquid scintillation. The percentage of total counts that the shifted complex represented was calculated for each sample. The graphed results are the average from several independent experiments, IMD: n = 3; CD1: n = 4; Rds2: n = 3. Error bars represent SEM. **(E)** WH domain mutations abolish DNA-binding activity of TAF1. TAF1-TAF7 complexes assembled in insect cells with WT-TAF1 or TAF1-WH mutant (WH3A) were analyzed by EMSA. Protein-bound and -unbound ^{32}P-labeled CD1 fragments were detected by autoradiography. **(F)** TAF1 mutations compromise growth complementation efficiency in *ts*13 cells. Mutant *ts*13 cells were transfected with the indicated CS2+ control or TAF1 expression plasmid at 33.5 °C and subsequently shifted to 39.5 °C. After 48-72 h at the elevated temperature, the number of viable cells was counted and expressed as a percent relative to cells transfected with WT-TAF1, given a value of 100%. The bar graph depicts the average from 4 independent complementation experiments. Error bars, SEM. CS2+, empty vector; WT, WT-TAF1; Δ848-850, deletion of residues 848-850; WH3A: R864/K865/K868 all changed to alanine; DPT-AAA: D976/P977/T978 all changed to alanine.

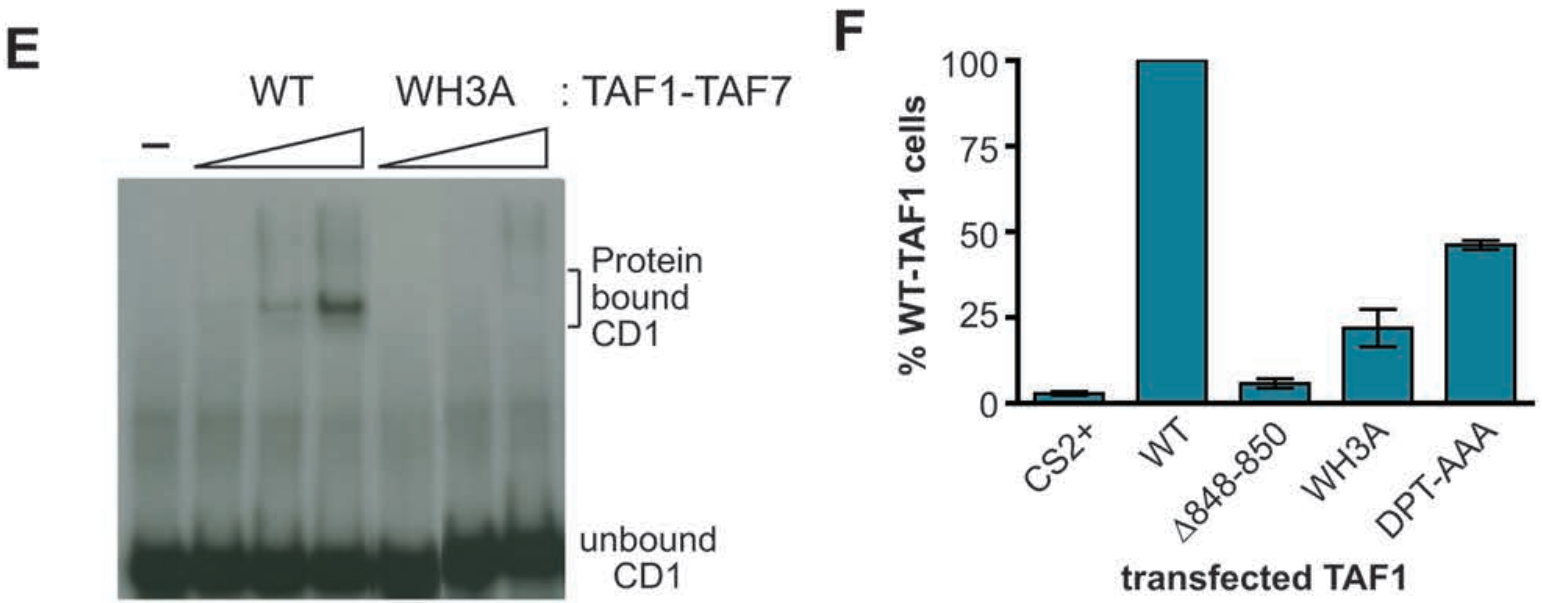


Figure 5 DNA binding and *ts*13 cell complementation activities of WT and mutant TAF1 proteins. **(A)** TAF1 displays DNA-binding activity. EMSA was carried out with increasing concentrations of the indicated purified proteins. Binding reactions were resolved on native 5% polyacrylamide and subjected to autoradiography. The positions of protein-bound and -unbound ^{32}P-labeled cyclin D1 (CD1) core promoter fragments are shown. **(B)** TAF1 binds to double-stranded DNA. TAF1-TAF7 (22.5 pmoles) was incubated with radiolabeled CD1 probes in the presence of the indicated molar excess amount of unlabeled double-stranded or single-stranded CD1. Reaction products were analyzed and detected as described in **A**. **(C)** Minimal sequence-specificity for TAF1 DNA binding. Increasing concentrations of TAF1-TAF7 complex was subjected to EMSA using different radiolabeled DNA probes. IMD: initiator and downstream region of super core promoter, CD1: cyclin D1 core promoter, Rds2: random double-stranded DNA sequence. Migrations of protein-bound and -unbound DNA fragments are indicated. **(D)** Quantitation of DNA binding to the DNA probes described in **C**. Regions containing the bound and unbound radiolabeled DNA probes were excised from the dried gel and quantified by liquid scintillation. The percentage of total counts that the shifted complex represented was calculated for each sample. The graphed results are the average from several independent experiments, IMD: n = 3; CD1: n = 4; Rds2: n = 3. Error bars represent SEM. **(E)** WH domain mutations abolish DNA-binding activity of TAF1. TAF1-TAF7 complexes assembled in insect cells with WT-TAF1 or TAF1-WH mutant (WH3A) were analyzed by EMSA. Protein-bound and -unbound ^{32}P-labeled CD1 fragments were detected by autoradiography. **(F)** TAF1 mutations compromise growth complementation efficiency in *ts*13 cells. Mutant *ts*13 cells were transfected with the indicated CS2+ control or TAF1 expression plasmid at 33.5 °C and subsequently shifted to 39.5 °C. After 48-72 h at the elevated temperature, the number of viable cells was counted and expressed as a percent relative to cells transfected with WT-TAF1, given a value of 100%. The bar graph depicts the average from 4 independent complementation experiments. Error bars, SEM. CS2+, empty vector; WT, WT-TAF1; Δ848-850, deletion of residues 848-850; WH3A: R864/K865/K868 all changed to alanine; DPT-AAA: D976/P977/T978 all changed to alanine.

Materials and Methods

Protein expression and purification

Human TAF1 (amino acids 600-1236) was co-expressed with the human TAF7 full-length protein as a glutathione S-transferase (GST) fusion protein in High Five (Invitrogen) insect cells. The TAF1-TAF7 complex was purified by glutathione affinity chromatography using lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0 mM DTT) supplemented with protease inhibitors. The protein complex was further purified by anion exchange and size exclusion chromatography (GE Healthcare) after off-column cleavage by the tobacco etch virus (TEV) protease. The eluted complex was concentrated by ultrafiltration to 12 mg/ml in buffer containing 25 mM Hepes, pH 7.5, 150 mM NaCl, 1.0 mM DTT. Trypsin (0.1% w/w) was added to the complex sample to trim off flexible regions during crystallization screening. The protein samples used for EMSA assays were purified the same way as described above except the final size exclusion was omitted.

Crystallization and data collection

The crystals of human TAF1-TAF7 complex were obtained at 4 °C by the hanging-drop vapor diffusion method, using 1.2 µl protein complex sample mixed with an equal volume of reservoir solution containing 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4, 20% w/v PEG3350. After being briefly equilibrated in the reservoir solution, the crystals were transferred to a cryoprotectant solution containing 25% w/v PEG3350, 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4 and 10% glycerol, and then flash-frozen in liquid nitrogen. The human TAF1-TAF7 complex heavy atom derivative crystals were prepared by soaking the native crystals in a buffer containing 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4, 20% w/v PEG3350 supplemented with 1.0 mM $K_2Pt(NO_2)_4$ for 12 h, followed by a back soaking in the reservoir solution for an additional hour. The crystals were subsequently harvested the same way as the native ones. All data sets were collected at the BL8.2.1 and BL8.2.2 beamlines at the Advanced Light Source of the Lawrence Berkeley National Laboratory. Native crystals ($P2_12_12_1$, a = 83.2 Å, b = 94.5 Å, c = 101.8 Å, $\alpha = \beta = \gamma = 90^\circ$) diffracted to 2.3 Å resolution, and heavy atom derivative crystals ($P2_12_12_1$, a = 83.1 Å, b = 94.5 Å, c = 102.2 Å, $\alpha = \beta = \gamma = 90^\circ$) diffracted to 2.8 Å resolution. X-ray diffraction data statistics are shown in Supplementary information, Table S1.

Structure determination

The human TAF1-TAF7 complex structure was determined by SAD phasing of a platinum protein derivative. Reflection data were indexed, integrated and scaled with the HKL2000 package [38]. Experimentally phased map with a well-defined solvent boundary and readily interpretable electron density for protein was calculated with PHENIX [39]. An initial model was established using PHENIX and refined against the native data set. The final model was manually built with the program COOT [40]. The complex structure was refined to an R_{factor} = 18.6% and an R_{free} = 23.4%. Crystallographic data statistics are shown in Supplementary information, Table S1. 99.81% of all residues of human TAF1-TAF7 complex are in the favored and allowed region of the Ramachandran plot. All structural figures were prepared using PyMOL (http://www.pymol.org/).

Gel filtration analysis

The TAF1-RAPiD (amino acids 1110-1236) and TAF7-CTD (amino acids 154-349) domains were expressed in *E. coli* in a modified pET-28a (Novagen) vector with a 6× His tag and a Protein GB1 tag, both cleavable by the TEV protease. Both proteins were purified by Ni-NTA affinity columns followed by overnight on-column TEV cleavage, and further by anion-exchange and size exclusion chromatography (GE Healthcare). To form complex, the two polypeptides were mixed in a 1:1 molar ratio, and incubated at 4 °C for one hour. Purified TAF1-RAPiD, TAF7-CTD and the formed complex were analyzed by gel filtration. The eluted complex fractions in the shifted peak were analyzed by 15% (w/v) SDS-PAGE gel and visualized by Coomassie-blue staining.

EMSA

TAF1-TAF7 complex (4-180 pmoles) purified from insect cells was incubated with 4 ng of ^{32}P 5′-end labeled DNA probes in 10 mM Hepes pH 7.9, 5 mM $MgCl_2$, 100 mM NaCl, 10% glycerol, 20 mM tetrasodium pyrophosphate, 0.2 mM dI:dC for 1 hour at 25 °C. For gel electrophoresis, 6x loading buffer (20% Ficoll, 0.025% bromophenol blue) was added and binding reactions were loaded onto nondenaturing 5% polyacrylamide (37.5:1 acrylamide:bis) that was pre-run in 0.5× TBE at 100 V for 30 min. Samples were resolved for 1.5 h at 100 V and shifted complexes detected by autoradiography. Competition assays using 22.5 pmoles of purified TAF1-TAF7 complex were performed as described above with the addition of either unlabeled CD1 core promoter double-stranded (20, 40, 80, 200 ng) or sense single-stranded (10, 20, 40, 100 ng) DNA to the binding reaction. DNA binding was quantified by excising the protein bound and unbound DNA bands from the dried gels. The amount of radioactivity in each band was measured by liquid scintillation (Tri-Carb B2810 TR, PerkinElmer). The percent DNA bound was calculated and plotted using Prism (GraphPad Software). The sequences of DNA probes using for EMSA are provided in Supplementary information, Table S2.

ts13 cell complementation assays

Mammalian TAF1 expression plasmids contain N-terminal HA-tagged TAF1 coding sequence inserted downstream of the CMV promoter in CS2+ vector. Point mutations in TAF1 were introduced by site-directed mutagenesis using the primers shown in Supplementary information, Table S2 and confirmed by DNA sequencing. *ts*13 cells were grown at 33.5 °C in Dulbecco's modified Eagles medium (Gibco) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and penicillin/streptomycin. For complementation assays, cells were seeded into 60 mm dishes, grown overnight to 70%-80% confluency, and transfected with 2.5 μg of CS2+ or TAF1 expression plasmids using FuGene HD transfection reagent (2:1 FuGene:DNA) according to the manufacturer's protocol (Roche). Transfected cells were maintained an additional 18-24 h at 33.5 C, after which half the cells were shifted to non-permissive temperature of 39.5 °C. The number of viable cells was determined after 36-48 h at 39.5 °C and the percentage relative to WT-TAF1 expression (given the value of 100%) was calculated.

Accession Code

Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 4RGW.