Molecular Cell, *Volume 39*

Supplemental Information

RNA Polymerase I Contains a TFIIF-Related DNA-Binding Subcomplex

Sebastian R. Geiger, Kristina Lorenzen, Amelie Schreieck, Patrizia Hanecker, Dirk Kostrewa, Albert J.R. Heck, and Patrick Cramer

SUPPLEMENTAL FIGURES

Figure S1. Alternative alignment of dimerization domain sequences from Pol I, II and III.

Structure-based alignment of amino acid sequences of *C. glabrata* A34.5 and A49 with their alternative counterparts in Pol II (*H. sapiens* Rap74 and Rap30, *S. cerevisiae* Tfg1 and Tfg2), and their alternative Pol III counterparts *S. cerevisiae* C37 and C53, respectively. All other labels are as in Figure 2E.

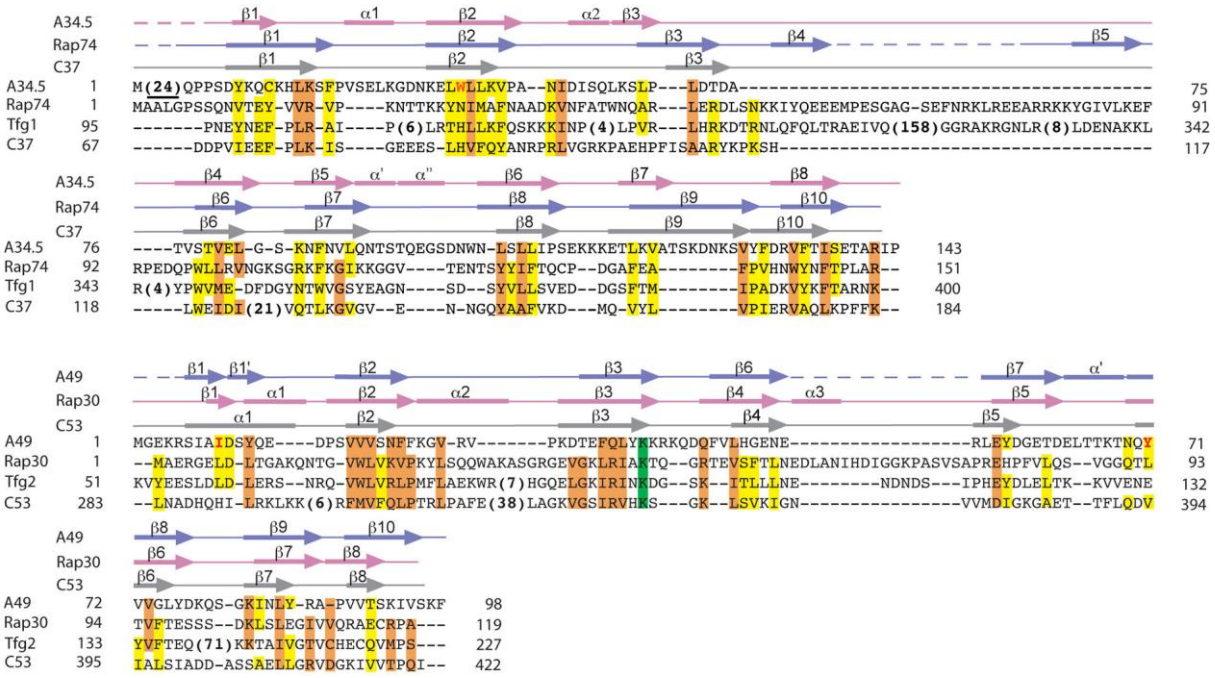


Figure S2. A49/34.5 tetramer formation

(A) Size-exclusion chromatography of the A49/34.5 dimerization module. Absorption at 280 nm and 260 nm is shown in blue and red, respectively. Molecular masses were determined by static light scattering. The protein was visualized by SDS-PAGE.

(B) Ribbon model of the A49/34.5 tetramer, composed of two A49/34.5 heterodimers. A49 is depicted in light blue and A34.5 in magenta. A pseudo 2-fold symmetry axis is indicated.

(C) β-strand exchange in the tetramer observed in crystals. View as in B. Secondary structure elements and A49 residues are indicated. A $2F_o$-F_c map is shown as a grey mesh and contoured at 1.0 sigma.

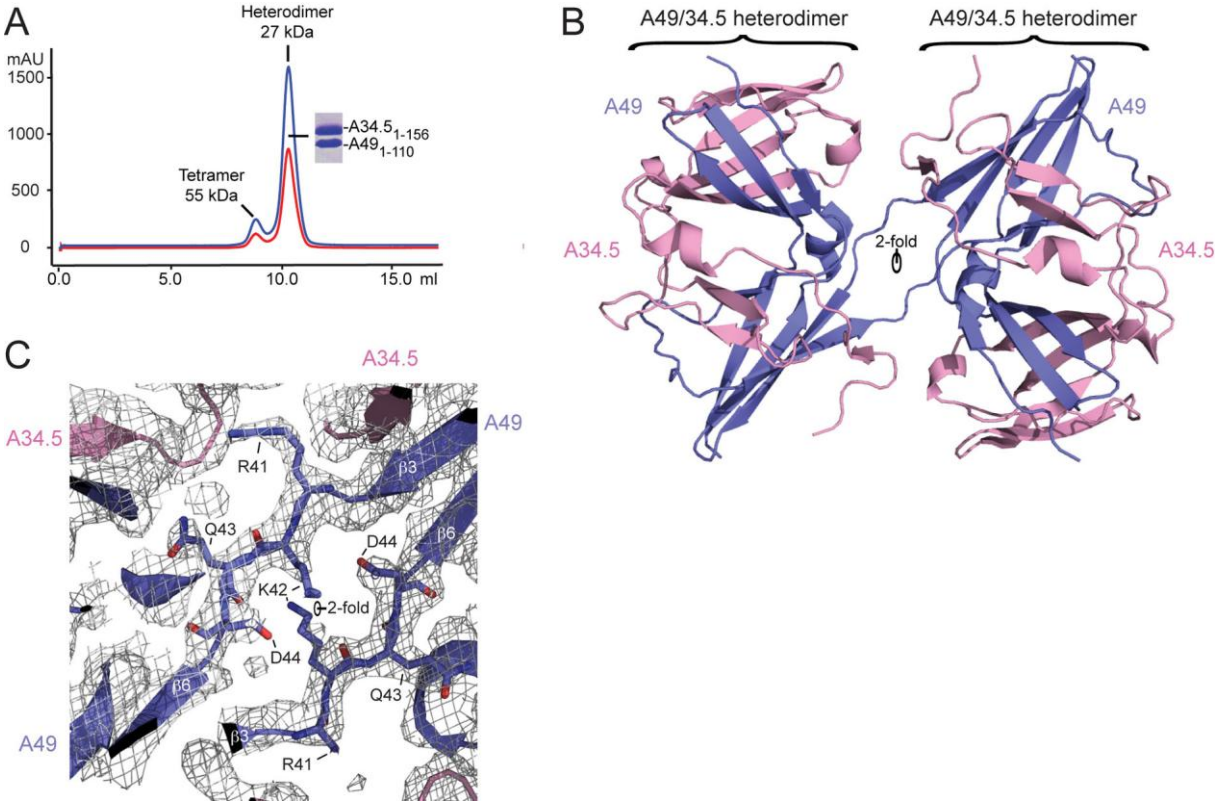


Figure S3. A49 tWH domain B-factor distribution and structure of the A49 linker domain

(A) Structure of the A49 tWH domain, colored according to B-factors. B-factor distribution is shown in rainbow colors ranging from blue (low B-factors) to red (high B-factors). Secondary structure elements with high B-factors are indicated.

(B) Structural superposition of the extended A49/34.5 dimerization module with the TFIIF Rap74/30 dimerization module structure (grey). Linker helices α' (A49) and α1 **(**Rap74) are in cyan.

(C) Observation of A49 linker helix α4 in one tWH molecule of the asymmetric unit of the structure in Fig. 3. The linker helix is in cyan, and neighboring domains in the crystal are in grey. Other colours are as in Fig. 2. The $2F_o$-F_c electron density map is contoured at 0.8σ (grey mesh).

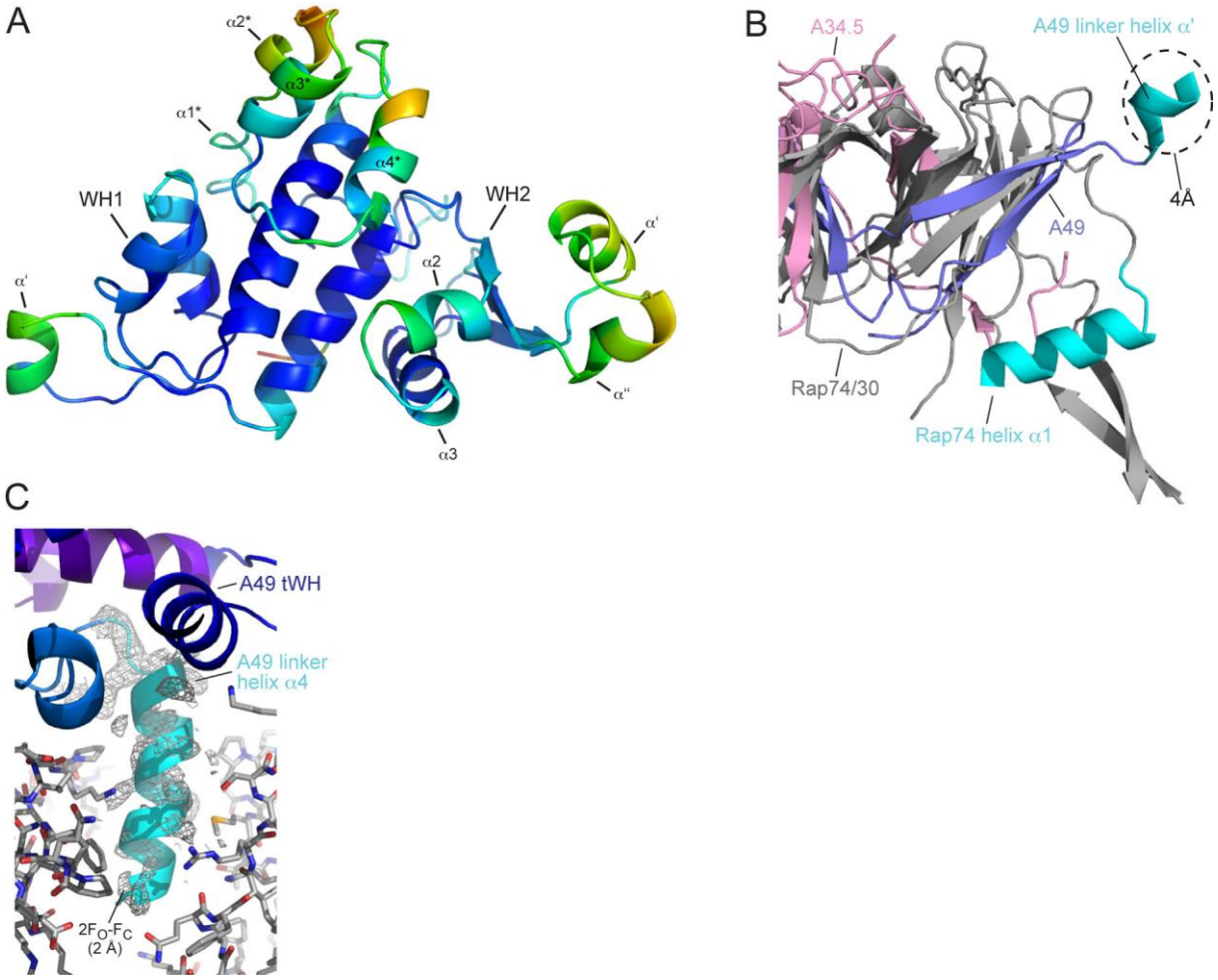


Figure S4. Homologous WH domains in Pol I/III subunits and Pol II-associated factors

Structure-based alignment of amino acid sequences of the *S. cerevisiae* A49 tWH domain (top) with their proposed human Pol II and Pol III counterparts TFIIE-β (PDB 1D8K (Okuda et al., 2004)) and C34 (PDB 2DK8 and 2DK5), respectively. Related individual WH domains in human TFIIF proteins Rap30 (PDB 1BBY (Groft et al., 1998)) and Rap74 (PDB 1I27 (Kamada et al., 2001)) are aligned for both WH1 and WH2 of A49.

For WH2 of TFIIEβ, predicted secondary structure elements are depicted in grey, when aligned by HHpred (p-value < 0.0001) (Soding et al., 2005) and predicted to be present by secondary structure propensity (Biegert et al., 2006). Residues predicted in the hydrophobic cores of individual WH domains are highlighted in black. Residues in Rap30 (Groft et al., 1998), TFIIEβ (Tanaka et al., 2009) and A49 (this study) that interact with dsDNA are in cyan. Additionally mutated A49 residues are in red. For TFIIEβ, mutated residues affecting *in vitro* transcription (Tanaka et al., 2009) are indicated with a 'T' below the sequence. For C34, mutated residues are colored in dark green when displaying a cryo-sensitive phenotype, or are indicated with numbers for a lethal hexa- (1) or double-mutant (2), respectively (Brun et al., 1997). A49 tWH subdomain interface residues are indicated with dots (hydrophobic) or triangles (hydrophilic), with matched colors indicating interactions. For TFIIEβ and C34, residues predicted to contribute to a putative tWH interface are indicated with grey dots. For TFIIEβ, grey triangles indicate two residues that are predicted to form an interface salt bridge. A residue in A49 and a residue in C34 that deviate in the structures from database entry sequences are marked with an asterisk below the protein sequence. All other labels are as in Figure 2E.

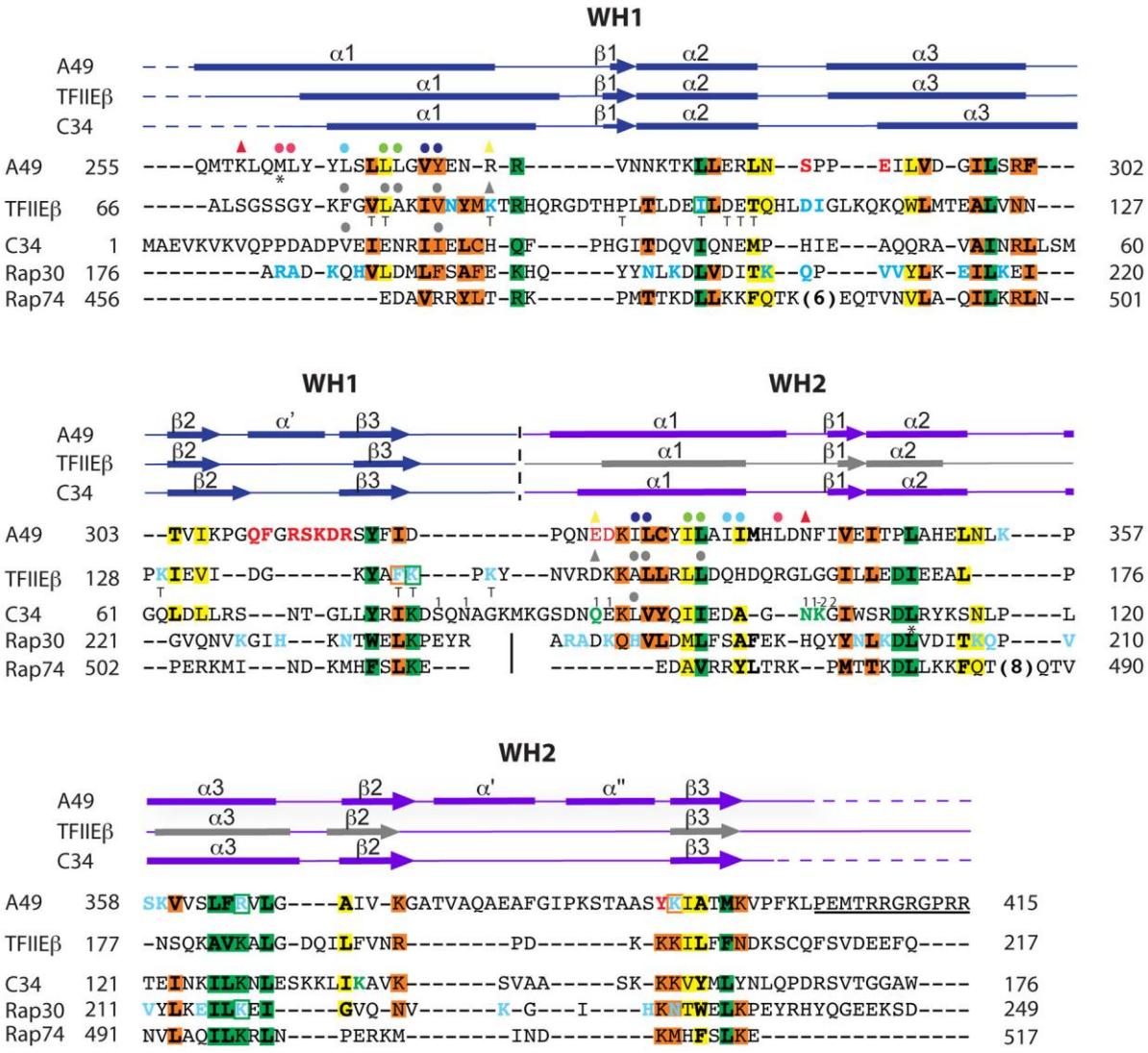

WH1
WH2
A49
TFIIEβ
C34
Rap30
Rap74

Figure S5. DNA binding of A49 tWH domain

(A) Modeled interaction between A49 WH1 and DNA. Putative DNA-interacting residues in A49 are shown as sticks and labeled.
(B) Modeled interaction between A49 WH2 and DNA.
(C) Comparison of two related A49 tWH structures. A49 $tWH_{183\text{-}403}$ is colored in blue (WH1) and purple (WH2), while A49 $tWH_{155\text{-}399}$ is in grey. WH1 helix α' is not ordered in the structure of A49 $tWH_{155\text{-}399}$, and parts of WH2 are shifted.
(D) Sequence alignment of the C-terminal tail from yeast and mammalian A49. Specific conserved basic residues and secondary structure elements are indicated.
(E) SDS PAGE analysis of various purified A49 tWH variants.

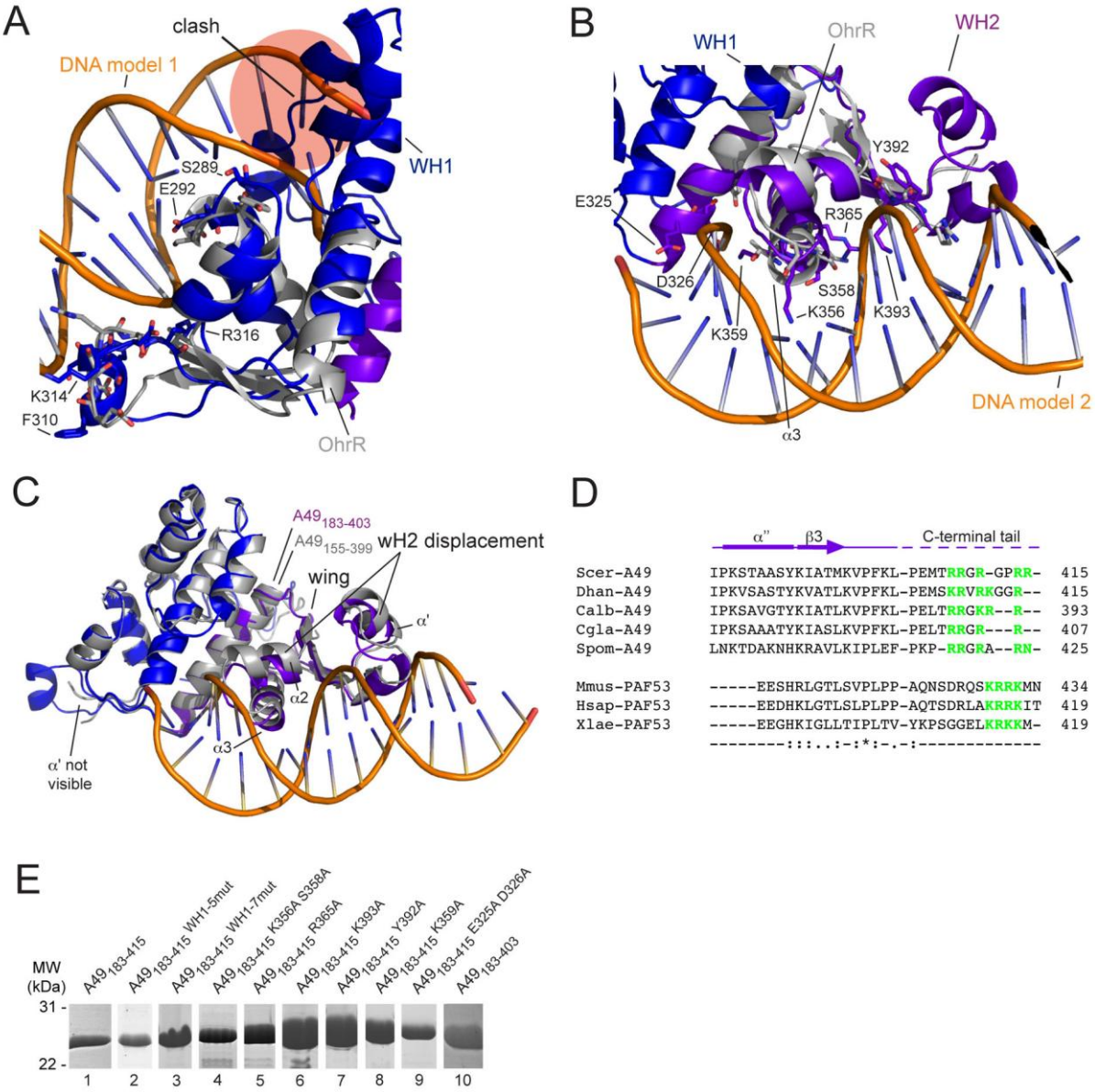


Figure S6. Additional interaction assay controls for A49/34.5 variants

Pulldown controls for various A49/34.5 variants. A49/34.5 variants are colored in red, degradation products are indicated with an asterisk. Wash and elution (elu) fractions are depicted above the corresponding lanes. $A49_{1\text{-}110}/34.5_{1\text{-}156}$ (lanes 1 and 2), $A49_{1\text{-}110}/34.5$ (lanes 3 and 4) and *C. glabrata* $A49_{1\text{-}190}/34.5_{1\text{-}156}$ (lanes 5 and 6).

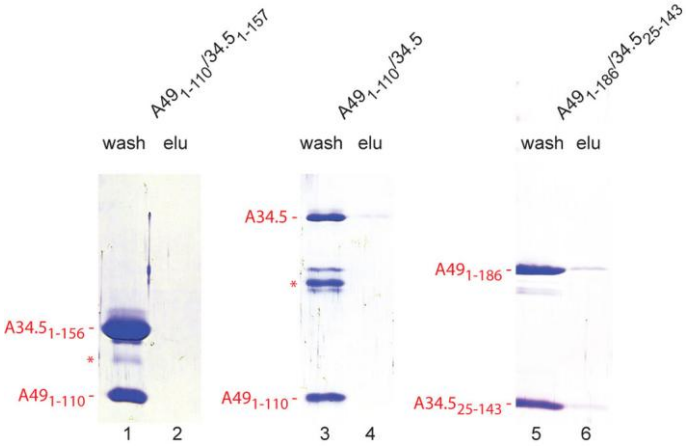
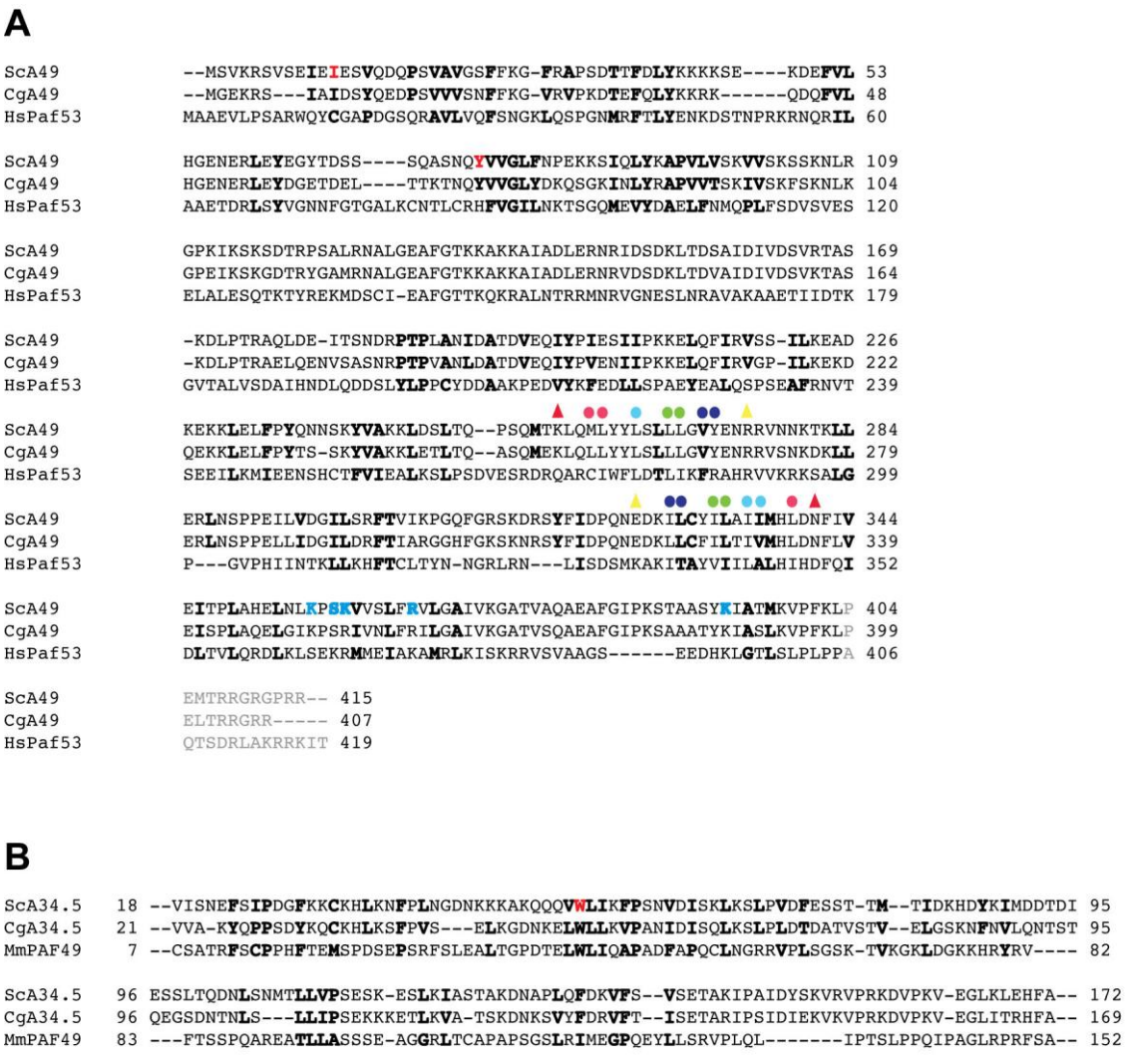


Figure S7. Sequence alignment of A49 and A34.5 with mammalian Paf53 and Paf49

Sequence alignment of amino acid sequences of **(A)** *S. cerevisiae* A49, *C. glabrata* A49 and *H. sapiens* Paf53, as well as **(B)** *S. cerevisiae* A34.5, *C. glabrata* A34.5 and *M. musculus* Paf49.

Residues present in the hydrophobic coresof the A49/34.5 dimerization module, the A49 tWH domain, and predicted in the mammalian counterparts, are highlighted in black. A49 residues that interact with dsDNA are in cyan, A49 and A34.5 residues that impair dimerization are in red. A49 tWH subdomain interface residues are indicated with dots (hydrophobic) or triangles (hydrophilic), with matched colors indicating interactions. Residues of the A49 C-terminal tail are colored in grey.



SUPPLEMENTAL TABLES

Supplemental table S1. A49 tWH homology statistics

	A49 WH1		A49 WH2	
	RMSD (Å)[1]	**Fold identity (%)[1]**	**RMSD (Å)[1]**	**Fold identity (%)[1]**
C34 WH1 (2DK8)[2]	**2.0**	**77.6**	2.9	59.0
C34 WH2 (2DK5)[2]	3.3	64.2	**2.4**	**60.2**
TFIIE-β WH1 (1D8K)[2]	**2.2**	**68.7**	3.7	61.5
TFIIE-β WH2 (modeled)[2]	3.0	62.4	**2.5**	**65.1**

[1]Deviations in Cα-positions (RMSD) and fold identities were analyzed by superposing the corresponding atomic structures with LSQKAB (Kabsch, 1976).

[2] PDB-codes are indicated. For TFIIEβ, the modeled structure (2D1H) was used. The homologous structural pairs are highlighted in bold and underlined letters.

Supplemental table S2. Nucleic acid sequences used in EMSA analysis

Name	Promoter position[1]	Label (5')[2]	Nucleotide sequence (5'-3', non-template strand) Reverse complementary sequence (3'-5', template strand)[3]
Non-sequence-specific dsDNA	- (47nt)	6-FAM -	5'-TGAGTAAGCCGTTTTGCAAGGAAGTGTGTTTCTTAAGTACTTGAGCT-3' 5'-AGCTCAAGTACTTAAGAAACACACTTCCTTGCAAAACGGCTTACTCA-3'
Ext. Upstream Dom. 2 (eUE-2)	-70 to -31	6-FAM -	5'-TAGTTTGTAATGGGAGGGGGGGTTTAGTCATGGAGTACAA-3' 5'-TTGTACTCCATGACTAAACCCCCCCTCCCATTACAAACTA-3'
Core Element linker (lCE)	-51 to -31	6-FAM -	5'-GGGTTTAGTCATGGAGTACAA-3' 5'-TTGTACTCCATGACTAAACCC-3'
Upstream Dom. 2 (UE-2)	-70 to -52	6-FAM -	5'-TAGTTTGTAATGGGAGGGG-3' 5'-CCCCTCCCATTACAAACTA-3'
Core Element Dom. 1 (CE-1)	-25 to +8	6-FAM -	5'-AGGAAAAGTAGTTGGGAGGTACTTCATGCGAAA-3' 5'-TTTCGCATGAAGTACCTCCCAACTACTTTTCCT-3'
Trun. Core Elem. (CE-1a)	-18 to +8	6-FAM -	5'-GTAGTTGGGAGGTACTTCATGCGAAA-3' 5'-TTTCGCATGAAGTACCTCCCAACTAC-3'
Trun. Core Elem. (CE-1b)	-30 to -1	6-FAM -	5'-GTGTGAGGAAAAGTAGTTGGGAGGTACTTC-3' 5'-GAAGTACCTCCCAACTACTTTTCCTCACAC-3'
ss - Upstream promoter (a)	-110 to -71	Cy5 -	5'-AGAGCGACAGAGAGGGCAAAAGAAAATAAAAGTAAGATTT-3' -
ss - Upstream promoter (b)	-100 to -61	Cy5 -	5'-AGAGGGCAAAAGAAAATAAAAGTAAGATTTTAGTTTGTAA-3' -
ss - Upstream promoter (c)	-90 to -51	Cy5 -	5'-AGAAAATAAAAGTAAGATTTTAGTTTGTAATGGGAGGGGG-3' -
ss - Upstream promoter (d)	-80 to -40	Cy5 -	5'-AGTAAGATTTTAGTTTGTAATGGGAGGGGGGGTTTAGTCAT-3' -
ss - Upstream promoter (e)	-74 to -35	Cy5 -	5'- GATTTTAGTTTGTAATGGGAGGGGGGGTTTAGTCATGGAG-3' -
ss - Upstream promoter (f)	-70 to -31	Cy5 -	5'-TAGTTTGTAATGGGAGGGGGGGTTTAGTCATGGAGTACAA-3' -
ss - Upstream promoter (g)	-65 to -26	Cy5 -	5'-TGTAATGGGAGGGGGGGTTTAGTCATGGAGTACAAGTGTG-3' -
ss - Upstream promoter (g)	-60 to -20	Cy5 -	5'-TGGGAGGGGGGGTTTAGTCATGGAGTACAAGTGTGAGGAAA-3' -
ss - Upstream promoter (h)	-49 to -10	Cy5 -	5'-GTTTAGTCATGGAGTACAAGTGTGAGGAAAAGTAGTTGGG-3' -
ss - Upstream prom. rev. (i)	-100 to -61	- Cy5	- 5'-TTACAAACTAAAATCTTACTTTTATTTTCTTTTGCCCTCT-3'
ss - Upstream prom. rev. (j)	-90 to -50	- Cy5	- 5'-CCCCCCTCCCATTACAAACTAAAATCTTACTTTTATTTTCT-3'
ss - Upstream prom. rev. (k)	-80 to -40	- Cy5	- 5'-ATGACTAAACCCCCCCTCCCATTACAAACTAAAATCTTACT-3'
ss - Upstream prom. rev. (l)	-70 to -31	- Cy5	- 5'-TTGTACTCCATGACTAAACCCCCCCTCCCATTACAAACTA-3'
ss - Upstream prom. rev. (m)	-60 to -20	- Cy5	- 5'-TTTCCTCACACTTGTACTCCATGACTAAACCCCCCCTCCCA-3'
ss - Upstream prom. rev. (n)	-49 to -10	- Cy5	- 5'-CCCAACTACTTTTCCTCACACTTGTACTCCATGACTAAAC-3'

[1] Pol I promoter positions are indicated according to the mapped (+1) initiation site (Moss et al., 2007).

[2] DNA-fragments are labeled at the 5'-end with either 6-FAM and Cy5 for double-stranded- or single-stranded DNA.

[3] Nucleotide sequence is shown for the non-template strand (5'-3') and for the template strand (3'-5') of the Pol I promoter.

Supplemental references

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallogr A *32*, 922-932.

Kamada, K., De Angelis, J., Roeder, R.G., and Burley, S.K. (2001). Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. Proc Natl Acad Sci U S A *98*, 3115-3120.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. Proteins *23*, 318-326.