Supporting Information

Radu, Schoenwetter et al.

**Production of the p34hs(1-233)/p44hs(321-395) complex.** The human p34(1-233) and p44(321-395) cDNAs were inserted into the pET28b (Novagen) and the pGEX (GE Healthcare) vectors using sequence and ligation independent cloning (SLIC) (1) and verified by sequencing. The p34hs(1-233) construct harbors an N-terminal cleavable hexa-histidine tag and p44hs(321-395) a GST-cleavable tag. Wild type and mutant (p34hs-R146E or p44hs-F374E) recombinant proteins were co-expressed in *E. coli* BL21 (DE3) cells. Cells were grown in LB medium at 37 °C and protein expression was induced with 0.5 mM isopropyl-β-thiogalactoside after an OD\textsubscript{600} of 0.7 – 0.8 was reached. Cells were further grown for 20 h at 18 °C, harvested and disrupted by sonication in buffer A (20 mM Tris-HCl pH 8, 250 mM NaCl and 1 mM DTT) containing 10 mM imidazole in the presence of an EDTA-free protease inhibitor cocktail (Complete™, Merck) and then clarified by centrifugation at 40,000 x g for 1 h at 4 °C. The clarified lysate was applied to a Ni-metal affinity resin column (5 ml His-Trap HP, GE Healthcare) and, after washing with buffer A containing increasing imidazole concentrations (10 mM, 25 mM and 75 mM), the bound His-p34hs(1-233)/GST-p44hs(321-395) complex was eluted in the same buffer containing 250 mM imidazole. Next, the complex was applied to a GST affinity resin column (Glutathione sepharose 4B, GE Healthcare). After extensive washing in buffer A, the His- and GST-tags were cleaved from both p34 and p44 using bovine thrombin. The complexes were finally purified by size exclusion chromatography (HiLoad 16/60 Superdex s200pg, GE Healthcare) (Figure S4, A and B) in buffer A containing 2 mM TCEP instead of DDT. The wild type p34hs/p44hs sample was further concentrated to 10 mg/ml using Vivaspin filtration units (Sartorius).

**Production of the p34ct(1-277)/p44hs(368-534) complex.** The genes encoding p34ct(1-277) and p44ct(368-534) were cloned from a cDNA library from *C. thermophilum* (provided by Ed Hurt). Using SLIC, p34ct(1-277) and p44ct(368-534) were cloned into the pBADM-11 vector (EMBL-Heidelberg) and the pETM-11 vector (EMBL-Heidelberg), respectively. p34ct(1-277) and p44ct(368-534) mutants were generated by site-directed mutagenesis and verified by sequencing (Eurofins Genomics or Seqlab). For co-expression of the wild-type proteins
p34ct(1-277) was expressed with an N-terminal hexa-histidine tag whereas p44ct(368-534) was expressed without any tag. For single protein expression the wild-type proteins as well as all variants were expressed with an N-terminal hexa-histidine tag. Protein expression was carried out in *E. coli* BL21-CodonPlus (DE3) RIL cells (Stratagene). Cells were grown in LB medium at 37 °C and protein expression was induced with 0.05% L-(+)-arabinose (for expression of the p34 variants) or 0.5 mM isopropyl-β-thiogalactoside (for expression of the p44 variants) or with both (for co-expression of wild-type p34 and p44) after an OD<sub>600</sub> of 0.6 – 0.8 was reached. Cells were further grown for 20 h at 15 °C, harvested and lysed in a buffer containing 20 mM Tris-HCl pH 7.5, 500 mM KCl and 0.5 mM TCEP using a mechanical cell disrupter (Microfluidics). After centrifugation at 38,000 x g for 1 h at 4 °C all proteins were purified using Ni-metal affinity chromatography (Ni-TED, Macherey-Nagel) followed by size exclusion chromatography (HiLoad 16/60 Superdex s200pg, GE Healthcare). Elution of the proteins during affinity chromatography was achieved with a buffer containing 20 mM Tris-HCl pH 8.0, 300 mM KCl, 250 mM imidazole and 1 mM TCEP. Size exclusion chromatography was carried out in either 20 mM CHES pH 9.5 (for co-purification of p34ct(1-277) and p44ct(368-534)) or 20 mM Tris-HCl pH 8.0 (for single purification of p34ct(1-277) or p44ct(368-534)), 150 mM KCl and 1 mM TCEP. The samples were concentrated to 5 – 10 mg/ml, based on their calculated extinction coefficients using ProtParam (SwissProt), via Amicon ultra centrifugal filters (Merck Millipore) and flash frozen in liquid nitrogen for storage at – 80 °C.

The ct sequences of p34 and p44 contain flexible linker insertions that are quite prominent in the ct proteins but are missing in the human proteins and have previously also been described for the p34ct vWA structure (1). The most prominent linker region in p44ct is reaching from residues 410 to 468 and is not visible in the electron-density map of the p34ct/p44ct minimal complex I. To improve crystal packing of the ct p34/p44 minimal complex this flexible linker region in p44ct(368-534) was replaced by five amino acids, namely S-N-G-N-G (ct p34/p44 minimal complex II). The artificial linker was introduced into p44ct(368-534) using SLIC and verified by sequencing. Protein expression and purification was performed as described above.
Structure determination and analysis.

**Determination of the p34ct(1-277)/p44ct(386-586) structure.** The ct p34/p44 minimal complex I was crystallized at 20°C in 15% (w/v) PEG 20,000 and 100 mM MES pH 6.5, whereas the ct p34/p44 minimal complex II (see Production of the p34ct(1-277)/p44hs(368-534) complex section for further information) was crystallized at 20°C in 5 – 10% (w/v) PEG 4,000, 20 – 33% (v/v) MPD and 100 mM HEPES pH 7.0 – 7.5. Crystals of the ct p34/p44 minimal complexes I and II were crystallized at protein concentrations of 5 – 10 mg/ml using the vapor diffusion method in sitting drops. Crystals of complex I used for structure solution took several weeks to grow but failed reproduction. Crystals of complex II were well reproducible and usually took five to ten days to appear. For data collection all crystals were washed in a drop containing the mother liquor supplemented with 20% (v/v) glycerol before they were flash frozen in liquid nitrogen. Data collection of complexes I and II was performed at 100 K and wavelengths of 0.8726 Å and 0.97625 Å at beamlines ID23-2 and BM14 (ESRF), respectively. The data sets were processed using either iMOSFLM and SCALA (2, 3) or XDS and AIMLESS (4, 5). Both complexes crystallized in space group P6$_3$22 with one copy of the heterodimer per asymmetric unit but varied in the unit cell dimensions (Table 1). As high resolution cutoff we have chosen a CC1/2 value of 0.6 for all data sets.

The structure of complex I was solved by molecular replacement via Phaser (6) using the structure of the p34ct vWA domain (PDB: 4PN7,(1)) as a search model. The p44ct RING domain was built into the electron density by using the human p44 NMR structure (PDB: 1Z60, (7)) and the weak anomalous signal of the two zinc ions as a guide. Model building and adjustment was performed in Coot (8). The final model of complex I was obtained using the higher resolution model of complex II after the latter had been fully refined. The structure of complex I was refined to a resolution of 3.7 Å with an R-factor of 19.6% and R$_{free}$ of 24.6% using REFMAC (9). This model contains 200 out of 277 residues of the p34ct vWA domain and 74 out of 167 residues of the p44ct RING domain. The structures of complex I and II do not show any significant differences (Figure S9) so that we exclusively discuss the higher resolution complex II model in this study.

The structure of complex II was solved by molecular replacement via Phaser using the structure of complex I as a search model. The structure was refined to a resolution of 2.2 Å
with an R-factor of 21.0% and R$_{\text{free}}$ of 22.7% using Phenix refine (10) and REFMAC5. The model was adjusted with Coot. The final model contains residues 18 – 89, 104 – 166 and 198 – 274 (212 out of 277 residues) of p34ct with residues 1 – 17, 90 – 103, 167 – 197 and 275 – 277 being presumably disordered. Taking into account that the artificial linker of p44ct could be modeled, the final model comprises residues 380 – 507 of p44ct (74 out of 113 residues in the improved model) with residues 368 – 379 and 508 – 534 being disordered.

**Determination of the p34hs(1-233)/p44hs(321-395) structure.** Crystals of the human p34(1-233)/p44(321-395) complex were washed for 30 s in a crystallization solution supplemented with 20% (v/v) PEG 400 or glycerol prior to freezing in liquid nitrogen. X-ray data were collected at 100 K using the PROXIMA1 beam line of the Soleil synchrotron facility (Gif-sur-Yvette, France) and the ID29 beamline of European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data were indexed and integrated using XDS (4); crystals belong to space group I2$_1$2$_1$2$_1$ (R$_{\text{merge}}$ of 13.3% and I B (Wilson) of 110 Å$^2$) with two complexes in the asymmetric unit (Table 1).

The structure was determined by molecular replacement using the p34ct vWA domain (PDB: 4PN7) as search model in the CCP4 software suite Phaser and the intrinsic Zn anomalous scattering signal of the two zinc ions in the p44 RING C-terminal domain was used to position this domain. The structure was extended and completed by iterative manual building in Coot (8) and refinement using Phenix (10) using the p34ct/p44ct structure as a guide (Table 1). The asymmetric unit contains two heterodimers: the model of the first complex is composed of 190 (out of 233) and 49 residues (out of 74) for p34hs and p44hs, respectively (chains A and B); the second complex is composed of 192 residues for p34 and 54 residues for p44 (chains C and D; see Figure 2B). Within each dimer, the relative orientation of the p34 and p44 subunits is identical to that observed in the complex from ct, which results in an overall RMSD of 1.5 Å for the 236 equivalent Cα (182 for p34 and 54 for p44).

Interfaces were analysed using PDBsum (11) and PISA (12).

**Characterization of protein complexes.**
*Isothermal titration calorimetry.* Thermodynamic parameters of molecular interactions between p34ct(1-277) and p44ct(368-534) were determined by isothermal titration calorimetry (ITC) using a VP-ITC instrument (MicroCal, GE Healthcare) at 37 °C and 260 rpm. All samples were degassed prior to the experiment and equilibrated in the same buffer conditions consisting of 20 mM CHES pH 9.5, 150 mM KCl and 1 mM TCEP. In all experiments, 40 – 50 μM wild-type or mutated p44ct(368-534) were titrated into the sample cell containing 4 – 5 μM wild-type or mutated p34ct(1-277). A volume of 10 μL was added at a time, resulting in 30 injections. The data were analyzed with a single-site binding model using the Origin software (OriginLab). All experiments were repeated at least twice. Control experiments were performed in which p44ct(368-534) was titrated into buffer.

*Size exclusion chromatography.* Analysis of the interaction between ct wild-type and mutant p34 and p44 proteins was also performed via analytical size exclusion chromatography (SEC) using a Superdex 200 10/300 GL or a Superdex 200 Increase 10/300 GL column (GE Healthcare). All SEC runs were performed in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP. 400 μL of a solution containing 12 μM protein was applied onto the column either alone or after mixing the two proteins in a 1:1 ratio and subsequent incubation on ice for 1 h. During sample application and elution at a flow rate of 0.5 mL/min the proteins were detected at 280 nm. Fractions of 0.5 mL were collected during the elution and analyzed via SDS-PAGE. Reference runs with wild-type proteins were always performed prior to runs with mutated proteins.

*Circular dichroism spectroscopy.* All measurements were performed in a 1 mm quartz cuvette employing a J-715 spectropolarimeter (Jasco) at wavelengths from 260 to 190 nm and room temperature. All samples were equilibrated in the same buffer conditions consisting of 20 mM K₂H₂PO₄ pH 8.1. To optimize the signal to noise ratio a total of 10 spectra were accumulated. The spectrum of the reference buffer was subtracted from each sample spectrum.

*p34hs/p44hs interaction assays.* Wild-type or mutated p34hs/p44hs complexes were expressed by infecting Sf21 insect cells (50 mL) with the appropriate baculoviruses and were resuspended in 25 mL of buffer A. Cell extracts were prepared as described above. Pull down
experiments were performed in buffer B (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.1% (v/v) Nonidet P-40, 1 mM TCEP) using 50 µL of Protein A Sepharose cross-linked with 1H5 anti-p44 antibody (directed against residues 1–17 of human p44) (Sigma) or StrepTactin Sepharose (IBA) for p34 pull downs. After extensive washing and equilibration in buffer B, proteins were eluted by competition with 2 column volumes (CV) of buffer B containing the appropriate synthetic peptide at 0.5 mg/mL for immunoprecipitations or 5 mM d-desthiobiotin (Sigma) for elution from StrepTactin Sepharose (IBA) and analyzed on a SDS-PAGE followed by Coomassie staining or by high resolution protein electrophoresis using the LabChip GXII System (Caliper, LifeSciences). The histograms represent the estimated ratios (in arbitrary units, au) between p34hs (wild-type and variants) and p44hs (wild-type and variants).

**Protein melting point analysis.** Protein thermal stability was measured utilizing a label-free fluorimetric analysis and the Prometheus NT.48 (NanoTemper Technologies). Briefly, the shift of the intrinsic tryptophan fluorescence of proteins upon temperature-induced unfolding was monitored by detecting the emission fluorescence at 330 and 350 nm. Thermal unfolding was performed in nanoDSF grade high-sensitivity glass capillaries (NanoTemper Technologies) at a heating rate of 1 °C per minute. Protein melting points ($T_m$) were calculated from the first derivative of the ratio of tryptophan emission intensities at 330 and 350 nm. Measurements were performed at constant concentrations of 50 nM of the purified complexes in 20 mM Tris-HCl pH 8, 250 mM NaCl and 1 mM DTT.

**References:**

1. Schmitt DR, Kuper J, Elias A, Kisker C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. PLoS One. 2014;9(7):e102389.
2. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):271-81.
3. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 1):72-82.
4. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32.
5. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 7):1204-14.
6. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74.
7. Kellenberger E, Dominguez C, Fribourg S, Wasielewski E, Moras D, Poterszman A, et al. Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. J Biol Chem. 2005;280(21):20785-92.
8. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501.
9. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMACS for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67(Pt 4):355-67.
10. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):213-21.
11. de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic Acids Res. 2014;42(Database issue):D292-6.
12. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774-97.
SI Figures and Tables

**Figure S1** Comparison of the p34 vWA/p44 RING domain with the UbcH5b~Ub/Ark2C RING complex.

**Figure S2:** Circular Dichroism analysis p34ct vWA and p44ct RING variants

**Figure S3:** Mutational analysis of the p34ct/p44ct interface

**Figure S4:** Mutational analysis of the p34hs(1-233)/p44hs(321-395) interface

**Figure S5:** Mutational analysis of the interface between full length p34hs/p44hs

**Figure S6:** Stability of the p34/p44 heterodimer and HDMX analysis

**Figure S7:** C4 domain model of p34hs

**Figure S8:** Influence of the p34hs C-terminal domain on the integrity of TFIIH.

**Figure S9:** Overall superposition of the p34ct/p44ct complex I and the p34ct/p44ct complex II

**Table S1.** Representative X-ray and NMR structures of TFIIH components

**Table S2.** Interface alignment of Human and Ct p34 residues

**Table S3.** Interface alignment of Human and Ct p44 residues

---

**Figure S1** Comparison of the p34 vWA/p44 RING domain with the UbcH5b~Ub/Ark2C RING complex. A superposition of the p44 RING (yellow) with the Ark2C E3 RING (plum) is shown on the left. The p34hs/p44hs minimal complex (p34hs vWA/p44hs RING in orange and yellow, respectively) and the UbcH5b~Ub/Ark2C RING complexes (E2/E3 RING in magenta and plum, respectively) are displayed side by side in the same orientation.
Figure S2: Circular Dichroism spectroscopy of p34ct(1-277) wild-type and variants (A) and p44ct(368-534) wild-type and variants (B).
Figure S3: Mutational analysis of the p34ct/p44ct interface from size exclusion chromatography and ITC of wild-type and mutant complexes. (A) SEC of p34ct(1-277) with p44ct(368-534) wild-type or variants. p34ct(1-277) (red) and p44ct(368-534) (blue) were analyzed separately and in a 1:1 stoichiometry (green) mixed prior to SEC. (B) Quantification of the interaction between wild-type or variant p34ct(1-277) and p44ct(368-534) by ITC. The thermodynamic parameters for the association of the wild-type protein domains are: $K_D=11 \text{ nM}$, $n=0.71$, $\Delta H=-25 \text{ kcal/mol}$ and $\Delta S=-45 \text{ cal/mol/degree}$. 
**Figure S4:** Mutational analysis of the p34hs(1-233)/p44hs(321-395) interface from size exclusion chromatography of wild-type and mutant complexes. The p34hs(1-233)/p44hs(321-395) wild type complex, the p34hs-R146E and the p44hs-F374E variants were co-expressed in *E. coli* and purified using nickel affinity chromatography followed by a GST pull-down. Eluates from the GST-affinity purification step were analysed via SDS-PAGE (A), loaded on a Superdex S200 16/60 gel filtration column (B and C) or subjected to a thermal stability analysis (D).
**Figure S5:** Mutational analysis of the interface between full length p34hs/p44hs. Full length wild-type or variants p34hs and p44hs were co-expressed in insect cells and the association of the two proteins was analysed using pull down experiments directed against p34hs (using strep-tag affinity) or p44hs (using the 1H5 monoclonal antibody directed against p44hs) (A) SDS-PAGE analysis followed by Coomassie staining of the pull-down experiments using p44hs as bait (IP p44hs) (B) Quantification of the pull down analysis using capillary electrophoresis (IP p34hs and IP p44hs). (C) Quantification of the NER and transcription activity of core-TFIH harbouring mutations in p34hs or p44hs from Figures 3E and 3F. Activities from independent experiments (n=3 and n=2, respectively) were normalized to wild-type and averaged.
**Figure S6:** Stability of the p34/p44 heterodimer and HDMX analysis. (A) Full length strep-tagged p34hs or p34hs(1-233) and p44hs were co-expressed in insect cells and the complex was purified using strep-tag affinity chromatography. Purified complexes (left panel) were analysed using the NanoTemper Technologies Prometheus instrument to evaluate their thermal stability (right panel). (B and C) Hydrogen-Deuterium eXchange coupled to Mass Spectrometry experiments in which the deuterium exchange labelling of p34hs full length in the presence and absence of p44hs was compared. The sequence coverage of p34hs is shown in Figure (B). Mapping of the measured backbone amide deuterium uptake after 30 sec, 1 min, 5 min and 10 min on the 3D structure of the p34hs vWA domain is illustrated in (C).
Figure S7: C4 domain model of p34hs (A) Sequence alignment of the C4 domains of p34ct and p34hs. The alignment was prepared using Clustal omega and colored with the ESPRIPT server. (B) Model of the C4 domain of p34hs. The structure is depicted in ribbon mode with a transparent surface. The area that shows the lowest deuterium exchange rate in the C4 domain is labeled in red and the residues are depicted in stick mode. The model was prepared using the Phyre2 server and manual corrections of the model were made in coot.
**Figure S8**: Influence of the p34hs C-terminal domain on the integrity of TFIIH. (A) Extracts from 200 $10^6$ Sf9 cells co-infected by viruses expressing core-TFIIH wild-type (rIIH6/p34 wt) or mutant (rIIH6/p34(1-233)) were purified using IMAC affinity chromatography followed by immunoprecipitation with an anti-p44 antibody. Purified proteins were resolved by SDS-PAGE followed by Coomassie staining where 5 µl of the eluate were analyzed (1/40 of total) (left panel) or by immune detection using specific antibodies where 0.1, 0.3, 1.0 and 3.0 µl of the eluate were loaded on the gel (right panel). (B) Extracts of SF9 cells co-infected by viruses expressing core-TFIIH wild-type (rIIH6/p34 wt) or mutant (rIIH6/p34(1-233)) were incubated with XPD plus CAK, immuno-precipitated using an antibody directed against cdk7 and immobilized proteins were analyzed using specific antibodies. Heavy (black circle) and light chains (black star) of the anti-cdk7 antibody are indicated. Lanes 1, 4, 5 and 8 have been used in Figure 4D.
**Figure S9:** Superposition of the p34ct/p44ct complex I (in dark blue and dark green for p34 and p44, respectively) and the p34ct/p44ct complex II (in cyan and light green for p34 and p44, respectively). The blue spheres represent the Zn ions.
## Table S1

| Structure          | Organism            | Method     | Resolution | Authors                      | Year   | PMID     |
|--------------------|---------------------|------------|------------|------------------------------|--------|----------|
| TFIIE                   | Human              | EM RTC     | 3.8Å       | Schultz et al., Cell 2000    | 2000   | 11007478 |
| TFIIE                   | Saccharomyces cerevisiae | EM 2D     | 3.5Å       | Chang et al., Cell 2000      | 2000   | 11007479 |
| TFIIE                   | Trypanosoma brucei | EM RTC     | 3.5Å       | Lee et al., NAR 2009         | 2009   | 19386623 |
| TFIIE                   | Saccharomyces cerevisiae | EM RTC     | 3.5Å       | Gibbons et al., PNAS 2011    | 2011   | 22308316 |
| Pol II PIC              | Saccharomyces cerevisiae | Cryo-EM   | 3.5Å       | Murakami et al., Science 2013 | 2013   | 24072820 |
| Pol II PIC              | Saccharomyces cerevisiae | Cryo-EM     | 3.5Å       | Robinson et al., Cell 2016   | 2016   | 27610567 |
| Pol II PIC              | Human              | Cryo-EM     | 3.5Å       | He et al., Nature 2012       | 2012   | 23446344 |
| Pol II PIC              | Human              | Cryo-EM     | 3.5Å       | He et al., Nature 2016       | 2016   | 27193682 |
| XPB                   | Human              | 4ERN       | 3.5Å       | Hilario et al Acta D, 2013   | 2013   | 23385459 |
| XPB                   | Archeoglobus fulgidus | 3WR       | 3.5Å       | Fan et al., Mol Cell, 2006   | 2006   | 16600867 |
| XPB                   | Archeoglobus fulgidus | 2Z4       | 3.5Å       | Fan et al., Mol Cell, 2006   | 2006   | 16600867 |
| XPB                   | Archeoglobus fulgidus | 2DL       | 3.5Å       | Fan et al., Mol Cell, 2006   | 2006   | 16600867 |
| p62                  | Human              | 1PEI       | 3.5Å       | Servais et al., 2004         | 2004   | 15195146 |
| p62                  | Human              | 2WNR       | 3.5Å       | Okuda et al., EMBO 2008      | 2008   | 18354501 |
| p62                  | Human              | 2RUK       | 3.5Å       | Okuda et al., JACS 2014      | 2014   | 25216154 |
| p62                  | Human              | 2DII       | 3.5Å       | Not published                |        | na       |
| p62                  | Saccharomyces cerevisiae | 2LOX     | 3.5Å       | Lafrance-Vanasse et al., NAR 2012 | 2012 | 22373916 |
| p62                  | Saccharomyces cerevisiae | 2M14     | 3.5Å       | Lafrance-Vanasse et al., NAR 2013 | 2013 | 23295669 |
| p62                  | Saccharomyces cerevisiae | 1VS0     | 3.5Å       | Di Lello et al., Biochemistry 2005 | 2005 | 15909982 |
| p62                  | Saccharomyces cerevisiae | 2GSO     | 3.5Å       | Di Lello et al., Mol Cell 2006 | 2006 | 16793543 |
| p62                  | Saccharomyces cerevisiae | 2K2U     | 3.5Å       | Langlois et al., JACS 2008   | 2008   | 18630911 |
| p62                  | Saccharomyces cerevisiae | 2L2I     | 3.5Å       | Mas et al., to be published  |        | na       |
| p62                  | Saccharomyces cerevisiae | 2MKR     | 3.5Å       | Chabot et al., Plos Path 2014 | 2014 | 24675874 |
| p8/p52               | Saccharomyces cerevisiae | 3DOM     | 3.5Å       | Kainov et al., NSBM 2008     | 2008   | 19172752 |
| p8/p52               | Saccharomyces cerevisiae | 3DGP     | 3.5Å       | Kainov et al., NSBM 2008     | 2008   | 19172752 |
| p34                  | Chaetomium thermophilum | 4PNZ     | 3.5Å       | Schmitt et al., PlosOne 2014 | 2014   | 25013903 |
| p34/p44               | Human              | na        | 3.5Å       | This work                    |        | na       |
| p44                  | Human              | 1Z60       | 3.5Å       | Kelleberger et al., 2006     | 2006   | 15790571 |
| p8                  | Human              | 2JNJ       | 3.5Å       | Vitorino et al., J Mol Biol 2007 | 2007 | 17350038 |
| p8                  | Human              | 1YDL       | 3.5Å       | na                          |        | na       |
| p8/p52               | Saccharomyces cerevisiae | 3DOM     | 3.5Å       | Kainov et al., NSBM 2008     | 2008   | 19172752 |
| p8/p52               | Saccharomyces cerevisiae | 3DGP     | 3.5Å       | Kainov et al., NSBM 2008     | 2008   | 19172752 |
| XPD                  | Thermoplasma acidophilum | 4A15     | 3.5Å       | Kuper et al., EMBO 2011      | 2011   | 22081108 |
| XPD                  | Thermoplasma acidophilum | 2VSF     | 3.5Å       | Kuper et al., Plos Biol 2008 | 2008 | 18576568 |
| XPD                  | Sulfolobus acidocaldarius | 3CRV    | 3.5Å       | Fan et al., Cell 2008        | 2008   | 18510924 |
| XPD                  | Sulfolobus tokodai | 2VL7       | 3.5Å       | Liu et al., Cell 2008        | 2008   | 18510925 |
| CDK7                 | Human              | 1UA2       | 3.5Å       | Loll et al, Structure 2004   | 2004   | 15530371 |
| cyclin H             | Human              | 1JKW       | 3.5Å       | Andersen et al., EMBO 1997   | 1997   | 9118957  |
| cyclin H             | Human              | 1KXU       | 3.5Å       | Kim et al., NSBM 1996        | 1996   | 8836101  |
| MAT1                 | Human              | 1Q25       | 3.5Å       | Gervais et al, JBC 2000      | 2000   | 11056162 |

Table S1. Published structures of TFIIE and complexes of subunits determined by EM, X-ray crystallography and NMR.
Table S2

| pH4Hs | HS | ASA (%) | BSA (%) | ALIGNED |
|-------|----|---------|---------|---------|
| LEU 57 | 0 | 0 | 0 | |
| ALA 58 | 0 | 0 | 0 | |
| VAL 59 | 0 | 0 | 0 | |
| LE 60 | 0 | 0 | 0 | |
| ALA 61 | 0 | 0 | 0 | |
| SER 62 | 0 | 0 | 0 | |
| PHE 63 | 0 | 0 | 0 | |
| LE 64 | 0 | 0 | 0 | |
| GLN 65 | 46 | 22 | 58 | POL |
| GLU 66 | 55 | 9 | 28 | CH |
| SER 67 | H | 7 | 5 | 78 | POL |
| ARG 68 | 40 | 2 | 10 | CH |
| PHE 69 | 47 | 14 | 86 | APOL |
| LEU 70 | 0 | 0 | 0 | |
| THR 71 | 37 | 0 | 63 | |
| PRO 72 | H | 20 | 3 | 26 | APOL |
| GLY 73 | 31 | 4 | 18 | APOL |
| LYS 74 | 117 | 89 | 18 | CH |
| GLY 75 | 125 | 6 | 10 | APOL |
| LEU 76 | 14 | 0 | 0 | |
| LYS 77 | 120 | 30 | 40 | |
| ASP 78 | 58 | 0 | 0 | |
| MET 79 | 0 | 0 | 0 | |
| THR 80 | 116 | 0 | 0 | |
| LYS 81 | 86 | 0 | 0 | |
| SER 82 | 37 | 0 | 0 | |
| ASP 83 | 86 | 0 | 0 | |
| LE 84 | 154 | 0 | 0 | |
| GLN 85 | 129 | 30 | 50 | POL |
| GLU 86 | 118 | 3 | 10 | CH |
| THR 87 | 87 | 0 | 0 | |
| GLU 88 | 48 | 0 | 0 | |
| THR 89 | 10 | 0 | 0 | |
| LEU 90 | 57 | 34 | 36 | APOL |
| LYS 91 | 9 | 0 | 0 | |
| ALA 92 | 11 | 2 | 88 | APOL |
| LYS 93 | 159 | 30 | 40 | APOL |
| SER 94 | 0 | 0 | 0 | |
| LEU 95 | 2 | 0 | 0 | |
| ALA 96 | 84 | 24 | 36 | APOL |
| LYS 97 | 76 | 65 | 19 | CH |
| GLU 98 | 49 | 0 | 0 | |
| LEU 99 | 34 | 0 | 0 | |
| ARG 100 | H | 96 | 78 | 20 | CH |
| ARG 101 | H | 89 | 59 | 30 | CH |
| MET 102 | 8 | 0 | 0 | |
| ASP 103 | 56 | 0 | 0 | |
| LYS 104 | 129 | 19 | 20 | CH |
| GLU 105 | 99 | 0 | 0 | |
| VAL 106 | 42 | 0 | 0 | |
| LYS 107 | 104 | 0 | 0 | |
| ASP 108 | 112 | 0 | 0 | |
| ASP 109 | 137 | 0 | 0 | |
| GLN 110 | 54 | 0 | 0 | |
| GLU 111 | 127 | 0 | 0 | |
| MET 112 | 26 | 0 | 0 | |
| LYS 113 | 83 | 0 | 0 | |
| SER 114 | 9 | 0 | 0 | |
| ARG 115 | 51 | 0 | 0 | |
| LEU 116 | 1 | 0 | 0 | |
| LEU 117 | 5 | 0 | 0 | |
| VAL 118 | 0 | 0 | 0 | |
| LE 119 | 2 | 0 | 0 | |
| LYS 120 | 0 | 0 | 0 | |
| ASP 121 | 0 | 0 | 0 | |
| LEU 122 | 0 | 0 | 0 | |
| LYS 123 | 70 | 19 | 20 | CH |
| HIS 124 | 69 | 0 | 0 | |
| ARG 125 | 42 | 0 | 0 | |
| THR 126 | 91 | 0 | 0 | |
| ASP 127 | 66 | 0 | 0 | |
| LYS 128 | 83 | 0 | 0 | |
| SER 129 | 83 | 0 | 0 | |
| ASP 130 | 84 | 0 | 0 | |
| THR 131 | 88 | 0 | 0 | |
| ARG 132 | 162 | 0 | 0 | |
| LYS 133 | 160 | 0 | 0 | |
| SER 134 | 102 | 0 | 0 | |
| GLY 135 | 41 | 0 | 0 | |
| LEU 136 | 34 | 0 | 0 | |
| ARG 137 | 115 | 0 | 0 | |
| HIS 138 | 94 | 0 | 0 | |
| LEU 139 | 32 | 0 | 0 | |
| THR 140 | 11 | 0 | 0 | |
| GLU 141 | H | 45 | 34 | 50 | POL |
| GLY 142 | H | 79 | 79 | 50 | POL |
| LEU 143 | 58 | 58 | 50 | APOL |
| LYS 144 | 126 | 28 | 19 | CH |
| LEU 145 | 1 | 0 | 0 | |
| SER 146 | 6 | 5 | 90 | |
| SER 147 | 123 | 22 | 78 | APOL |
| LYS 148 | 1 | 0 | 0 | |
| THR 149 | 1 | 0 | 0 | |
| LYS 150 | 28 | 18 | 80 | APOL |
| ARG 151 | 58 | 58 | 100 | APOL |
| HIS 152 | 44 | 26 | 74 | CH |
| HIS 153 | 6 | 0 | 0 | |
| LEU 154 | 115 | 0 | 0 | |
| ARG 155 | 15 | 0 | 0 | |
| LEU 156 | 13 | 0 | 0 | |
| ARG 157 | 90 | 0 | 0 | |
| LEU 158 | 108 | 28 | 30 | APOL |
| SER 159 | 99 | 0 | 0 | |
| LEU 160 | 42 | 0 | 0 | |
| THR 161 | 91 | 0 | 0 | |
| ASP 162 | 66 | 0 | 0 | |
| ASP 163 | 83 | 0 | 0 | |
| ARG 164 | 84 | 0 | 0 | |
| THR 165 | 88 | 0 | 0 | |
| ALA 166 | 162 | 0 | 0 | |
| ALA 167 | 102 | 0 | 0 | |
| GLY 168 | 41 | 0 | 0 | |
| LEU 169 | 34 | 0 | 0 | |
| HIS 170 | 115 | 0 | 0 | |
| ARG 171 | 15 | 0 | 0 | |
| ASP 172 | 90 | 0 | 0 | |
| LEU 173 | 5 | 0 | 0 | |
| ASP 174 | 3 | 0 | 0 | |
| ASP 175 | 69 | 0 | 0 | |
| ASP 176 | 84 | 0 | 0 | |
| ASP 177 | 91 | 0 | 0 | |
| GLN 178 | H | 49 | 22 | 50 | POL |
| THR 179 | 93 | 0 | 0 | |
| LE 180 | 126 | 5 | 10 | APOL |
| PRO 181 | H | 48 | 34 | 60 | POL |
| THR 182 | 1 | 0 | 0 | |
| THR 183 | 20 | 0 | 0 | |
| ARG 184 | H | 72 | 7 | 88 | APOL |
| ARG 185 | H | 6 | 6 | 100 | APOL |
| VAL 186 | 51 | 0 | 0 | |
| PHE 187 | 111 | 0 | 0 | |
| ASP 188 | 47 | 44 | 56 | APOL |
| LYS 189 | 27 | 0 | 0 | |
| HIS 190 | 261 | 23 | 20 | POL |
| HIS 191 | 57 | 57 | 100 | APOL |
| ARG 192 | 216 | 0 | 0 | |
| LE 193 | 12 | 0 | 0 | |
Tables S2 and S3. Structural alignment of human and Ct p34 vWA domains and p44 RING fingers. Interfaces were analysed with PISA (http://pdbe.org/pisa/). HS (residues making Hydrogen bonds or Salt bridges), ASA (Accessible Surface Area, Å²), BSA (Buried Surface Area, Å² and %). Residues in the Table are color coded: Interfacial residues in gold or yellow (according to their BSA, above or below 50%), buried in light blue and exposed in grey. Aligned interface residues are indicated by an orange or a red box (when identical).