nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-018-0156-z

Chaperonin CCT checkpoint function in basal transcription factor TFIID assembly

Simona V. Antonova[1,15], Matthias Haffke[2,13,15], Eleonora Corradini[3], Mykolas Mikuciunas[1], Teck Y. Low[3,14], Luca Signor[4], Robert M. van Es[5], Kapil Gupta[6], Elisabeth Scheer[7,8,9,10], Harmjan R. Vos[5], László Tora[7,8,9,10], Albert J. R. Heck[3], H. T. Marc Timmers[1,11,12*] and Imre Berger[6*]

TFIID is a cornerstone of eukaryotic gene regulation. Distinct TFIID complexes with unique subunit compositions exist and several TFIID subunits are shared with other complexes, thereby conveying precise cellular control of subunit allocation and functional assembly of this essential transcription factor. However, the molecular mechanisms that underlie the regulation of TFIID remain poorly understood. Here we use quantitative proteomics to examine TFIID submodules and assembly mechanisms in human cells. Structural and mutational analysis of the cytoplasmic TAF5–TAF6–TAF9 submodule identified novel interactions that are crucial for TFIID integrity and for allocation of TAF9 to TFIID or the Spt-Ada-Gcn5 acetyltransferase (SAGA) co-activator complex. We discover a key checkpoint function for the chaperonin CCT, which specifically associates with nascent TAF5 for subsequent handover to TAF6–TAF9 and ultimate holo-TFIID formation. Our findings illustrate at the molecular level how multisubunit complexes are generated within the cell via mechanisms that involve checkpoint decisions facilitated by a chaperone.

Eukaryotic gene transcription by RNA polymerase II (pol II) is controlled by a plethora of proteins that are preassembled into complexes, which include basal transcription factors[1,2]. TFIID is the first transcription factor to bind the core promoter, nucleating the pre-initiation complex. In humans, TFIID comprises the TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs). Cryo-electron microscopy (cryo-EM) analyses have provided important insights into TFIID architecture and promoter binding[3,4]. Distinct TFIID complexes have been identified, which contain paralogues of TAFs and TBP, that have key roles in cellular differentiation[1,2,5]. The human TFIID subunits TAF9, TAF10 and TAF12 are also present in SAGA[6], a co-activator complex that, similar to TFIID, is globally required for efficient pol II transcription[7,8]. In yeast, Taf5 and Taf6 are also present in the SAGA complex, whereas the human SAGA complex contains the paralogues TAF5L and TAF6L[9]. Sharing of subunits is not unique to TFIID and SAGA complexes but occurs frequently in complexes that control transcription and chromatin conformation[1,2,6–11]. Faithful subunit allocation and regulated complex assembly therefore must be essential for proper cell development and function; however, the underlying molecular mechanisms and factors that are involved remain unknown. Chaperones are known to facilitate the folding of key proteins in essential processes, such as kinase-mediated cellular signalling, proteasome assembly or nucleosome turnover[11–15]. The CCT complex has been shown to mediate ATP-dependent folding of many substrates, including actin, tubulin and the recently described CRL^{CSA} DNA-repair factor[16–18]. CCT adopts a barrel-like shape, with two hetero-octameric rings stacked on top of each other creating two folding chambers[18].

Here we use targeted quantitative mass spectrometry (qMS)[19,20] to systematically analyze TAF interactions in the nucleus and cytoplasm of human cells in order to investigate the molecular mechanisms underlying TFIID formation. We identified a cytoplasmic TFIID submodule that comprises TAF5, TAF6 and TAF9. The crystal structure of this complex revealed complex TAF–TAF interactions, providing a basis for mutational studies. Notably, our analyses reveal a key checkpoint function for CCT, providing unprecedented insights into the early steps of holo-TFIID assembly, which is regulated by this chaperonin.

Results

TFIID and SAGA submodules in the cytoplasm. The presence of stable, partial TFIID assemblies in cells suggests that the holo-complex occurs is formed by discrete, functional submodules[3,21–23]. TAF5 is the presumed central scaffold within TFIID and has been shown to interact with TAF6 and TAF9[3,24]. TAF6 and TAF9 form a heterodimer that is stabilized by the pairing of their histone-fold domains (HFDs)[25]. We created stable doxycycline (Dox)-inducible HeLa cell lines[26] that express TAF5, TAF6 or TAF9 proteins that contained an N-terminal green-fluorescent protein (GFP) tag. As expected, all GFP–TAF proteins predominantly localized to the nucleus

[1]Molecular Cancer Research and Regenerative Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands. [2]European Molecular Biology Laboratory (EMBL), Grenoble, France. [3]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences and Netherlands Proteomics Centre, Utrecht University, Utrecht, The Netherlands. [4]Université Grenoble Alpes, CEA, CNRS, IBS (Institut de Biologie Structurale), Grenoble, France. [5]Molecular Cancer Research, Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands. [6]Bristol Synthetic Biology Centre BrisSynBio, Biomedical Sciences, School of Biochemistry, University of Bristol, Bristol, UK. [7]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France. [8]Centre National de la Recherche Scientifique, UMR 7104, Illkirch, France. [9]Institut National de la Santé et de la Recherche Médicale, U964, Illkirch, France. [10]Université de Strasbourg, Illkirch, France. [11]Department of Urology, Medical Center–University of Freiburg, Freiburg, Germany. [12]Deutsches Konsortium für Translationale Krebsforschung (DKTK), Standort Freiburg, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany. [13]Present address: Chemical Biology and Therapeutics, Novartis Institutes for BioMedical Research, Basel, Switzerland. [14]Present address: UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia. [15]These authors contributed equally: Simona V. Antonova, Matthias Haffke. *e-mail: m.timmers@dkfz-heidelberg.de; imre.berger@bristol.ac.uk

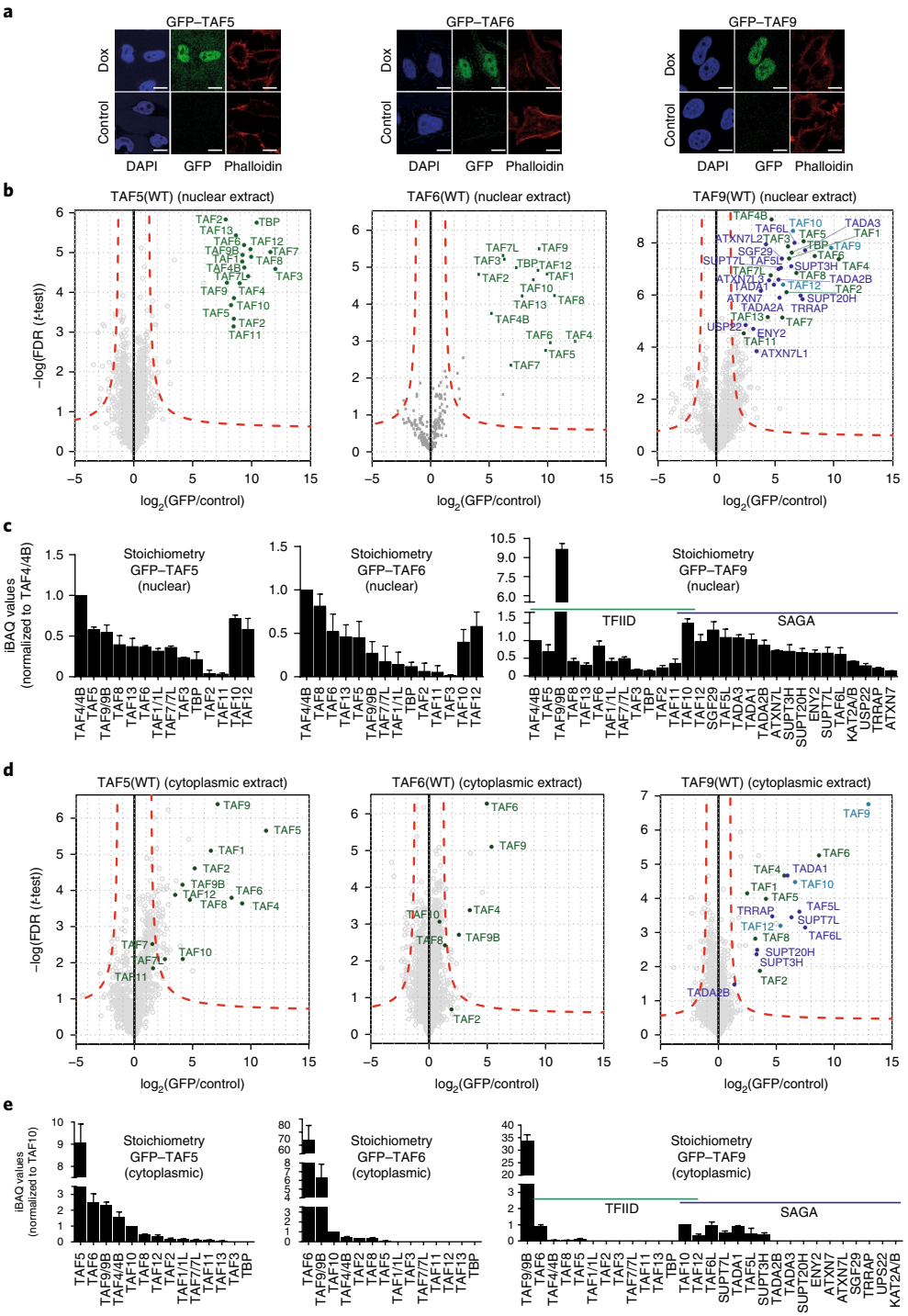

Fig. 1 | TFIID and SAGA submodules in the cytoplasm. a, Cellular localization of GFP–TAF5, GFP–TAF9 and GFP–TAF6 by confocal microscopy. Scale bars,10 μm. **b**, GFP–TAF5 and GFP–TAF6 enrich all TFIID subunits in co-IPs from nuclear extracts; GFP–TAF9 enriches TFIID and SAGA subunits. Subunits unique to TFIID are colored in green, subunits unique to SAGA are shown in purple, shared subunits are shown in blue. Dashed red lines denote the threshold between background and significant enrichment (two-tailed Student's *t*-test; FDR = 1%; $S_0 = 1$). Each data point is plotted as the mean of technical triplicates. **c**, Relative abundance of TFIID and SAGA subunits from nuclear extracts are shown, normalized to TAF4/4B. Data are mean ± s.d. of technical triplicates. **d**, GFP–TAF5 and GFP–TAF6 enrich a subset of TFIID subunits in co-IPs from the cytoplasm; GFP–TAF9 enriches both TFIID and SAGA subunits. Each data point is plotted as mean of technical triplicates. **e**, Relative abundance of TFIID and SAGA subunits in the cytoplasmic co-IPs. Data are mean ± s.d. of technical triplicates. Source data for **c** and **e** are available in the online version of the paper.

(Fig. 1a). GFP–TAF5 expression levels were close to endogenous levels, whereas GFP–TAF6 expression was increased. Anti-GFP immunoblotting and quantitative real time PCR (qPCR) showed that GFP–TAF9 levels were between GFP–TAF5 and GFP–TAF6 expression levels (Supplementary Fig. 1). We examined integration of GFP–TAF proteins into TFIID by GFP co-immunoprecipitation (co-IP) from nuclear extracts of Dox-treated cells followed by intensity-based absolute quantification (iBAQ) qMS[27]. All GFP–TAF

Table 1 | Data collection, phasing and refinement statistics

	Native[a]	Ta_6Br_{12}[a]
Data collection		
Space group	I23	I23
Cell dimensions		
a, *b*, *c* (Å)	339.05, 339.05, 339.05	338.16, 338.16, 338.16
α, *β*, *γ* (°)	90, 90, 90	90, 90, 90
		Peak
Wavelength	1.0044	1.25439
Resolution (Å)[b]	79.92–2.50 (2.60–2.50)[b]	48.81–3.80 (3.90–3.80)[b]
R_{merge}	0.1358 (3.048)	0.173 (1.17)
$I/\sigma(I)$	8.67 (0.64)	9.48 (2.08)
$CC_{1/2}$	0.995 (0.188)	0.996 (0.676)
Completeness (%)	99.03 (91.70)	100.00 (100.00)
Redundancy	4.69 (4.15)	5.78 (5.70)
Refinement		
Resolution (Å)	79.92–2.50	
Number of reflections	219,090	
R_{work}/R_{free}	19.73/22.00	
Number of atoms		
Protein	24,768	
Ion (Cl^-)	7	
Water	391	
B factors		
Protein	88.6	
Ligand/ion	79.6	
Water	68.8	
Root-mean-square deviations		
Bond lengths (Å)	0.004	
Bond angles (°)	1.06	

[a]A single crystal was used to collect the native (PDB 6F3T) and the Ta_6Br_{12} dataset. [b]Values in parentheses are for the highest-resolution shell.

proteins were enriched in the complete holo-TFIID complex (Fig. 1b,c). TAF9 is shared between TFIID and SAGA[6], and all SAGA subunits were similarly enriched in nuclear GFP–TAF9 co-IPs. We determined the relative abundance of TFIID subunits and found similar stoichiometries in all experiments. Moreover, comparable levels of SAGA and TFIID were present in the GFP–TAF9 co-IPs (Fig. 1c). Our results demonstrate that GFP-tagged TAF5 and TAF6 are efficiently integrated into TFIID, whereas GFP–TAF9 is found in both TFIID and SAGA complexes.

Next, we analyzed GFP–TAF proteins that were purified from the corresponding cytoplasmic extracts (Fig. 1d,e). We found that TAF6 and TAF9 co-purified efficiently with GFP–TAF5, revealing the presence of a stable heterotrimeric complex in the cytoplasm. We also found significant levels of TAF4, TAF8, TAF10 and TAF12, similar to stable TFIID subassemblies in recombinant reconstitution experiments[3]. Cytoplasmic GFP–TAF6 predominantly co-purified TAF9. Cytoplasmic GFP–TAF9 co-IPs showed enrichment of TAF6 as well as the SAGA subunits TAF5L, TAF6L, TADA1, SUPT3H and SUPT7L. TAF10 and TAF12, like TAF9, are shared between TFIID and SAGA, and were also present. In SAGA, TAF10 and TAF12 are known to form specific heterodimers with SUPT7L and TADA1, respectively. Therefore, the SAGA module that we observed may represent a putative core-SAGA complex that is similar to the core-TFIID complex that was originally identified in *Drosophila melanogaster*[3,21]. Taken together, our data provide evidence for the presence of discrete TFIID and SAGA submodules in the cytoplasm, which likely represent building blocks of the respective holo-complexes.

Crystal structure of TAF5–TAF6–TAF9. Our qMS experiments revealed TAF5–TAF6–TAF9 as a prevalent TFIID submodule in the cytoplasm. Limited proteolysis of full-length TAF5–TAF6–TAF9 identified a sample that was suitable for crystallization (Supplementary Figs. 2, 3 and Supplementary Table 1). Crystals comprising a TAF5 construct, which spanned the N-terminal domain (NTD) and WD40-repeat domain (TAF5(194–800)), bound to an extended TAF6–TAF9 heterodimer (TAF6(1–92); TAF9(1–120)) diffracted to high resolution (Supplementary Fig. 4). Phases were obtained with tantalum bromide clusters and the structure was refined to 2.5 Å (Table 1), resulting in excellent electron densities (Supplementary Fig. 4e–h). The final model includes residues 207–800 of TAF5 (residues 194–206, 381–416 and 748–753 were not modeled owing to poor definition), residues 6–92 of TAF6 and residues 5–120 of TAF9.

The crystal structure displays a compact, triangular shape (Fig. 2a), with the TAF6–TAF9 HFD heterodimer sandwiched between the NTD and WD40-repeat domain of TAF5, giving rise to an intertwined architecture that has TAF5 as the central scaffold. The TAF5 NTD was crystallized previously in isolation[28]. Substantial differences are apparent in the TAF5–TAF6–TAF9 complex. Notably, the N-terminal α-helix 7 of the NTD is rearranged owing to intimate interactions with the TAF5 WD40-repeat domain (Supplementary Fig. 5). The WD40 repeat adopts a seven-bladed β-propeller with pronounced α-helical insertions in blades 1 and 7. TAF5 engages extensively with the central α-helix of the TAF6 HFD through its NTD and the bottom cavity of the WD40 domain (Fig. 2b). The HFDs of TAF6 and TAF9 adopt a structure similar to the *D. melanogaster* TAF6–TAF9 HFD pair[29].

Our human TAF9 construct comprises an extended C-terminal region, which is completely resolved in the electron density (Fig. 2c). This TAF9 C-terminal domain adopts a distinct conformation that is crucial for TAF5–TAF6–TAF9 complex integrity, wrapping around the TAF5 WD40-repeat domain like a clamp with three major anchor points (Fig. 2d). The TAF9 C-terminal loop region attaches tightly to the surface of the TAF5 WD40 domain, involving residues L104, I107 and L115 (Fig. 2e). The second anchor is formed by the TAF9 αC helix, which packs laterally against the TAF5 WD40-repeat domain, engaging in multiple hydrogen bonds that are mediated by R99, N100 and T102 (Fig. 2f). Moreover, a triple proline turn formed by TAF9 residues P86–P88 interacts with both the TAF5 NTD and WD40-repeat domain (Fig. 2g). This remarkable multitude of protein–protein interactions combine to a total buried surface area of 2,713 $Å^2$ in the TAF5–TAF6–TAF9 complex.

TAF5–TAF9 interactions dictate TFIID assembly. Our crystal structure revealed unexpected, complex interactions with TAF9, which we next analyzed in cells by qMS, using a series of mutants (Fig. 2h). Four TAF9 mutants (TAF9m1–m4) were designed with the aim to disrupt the TAF5–TAF9 interface, while maintaining correct folding of the proteins. TAF9m1 targeted the hydrophobic side chains of the α-helix that are inserted into the hydrophobic pockets on the TAF5 WD40 surface as well as prominent salt bridges in between the TAFs. In TAF9m2, we abolished the network of hydrogen bonds between TAF5 and TAF9. TAF9m3 was designed to destabilize the triple proline turn in the C-terminal extension and abolish the $TAF5^{S303}$–$TAF9^{R89}$ interaction. Finally, TAF9m4 combined sall of these mutations.

The GFP–TAF9 mutants were expressed in HeLa cells and co-IPs of nuclear extracts were subjected to qMS (Fig. 3a and

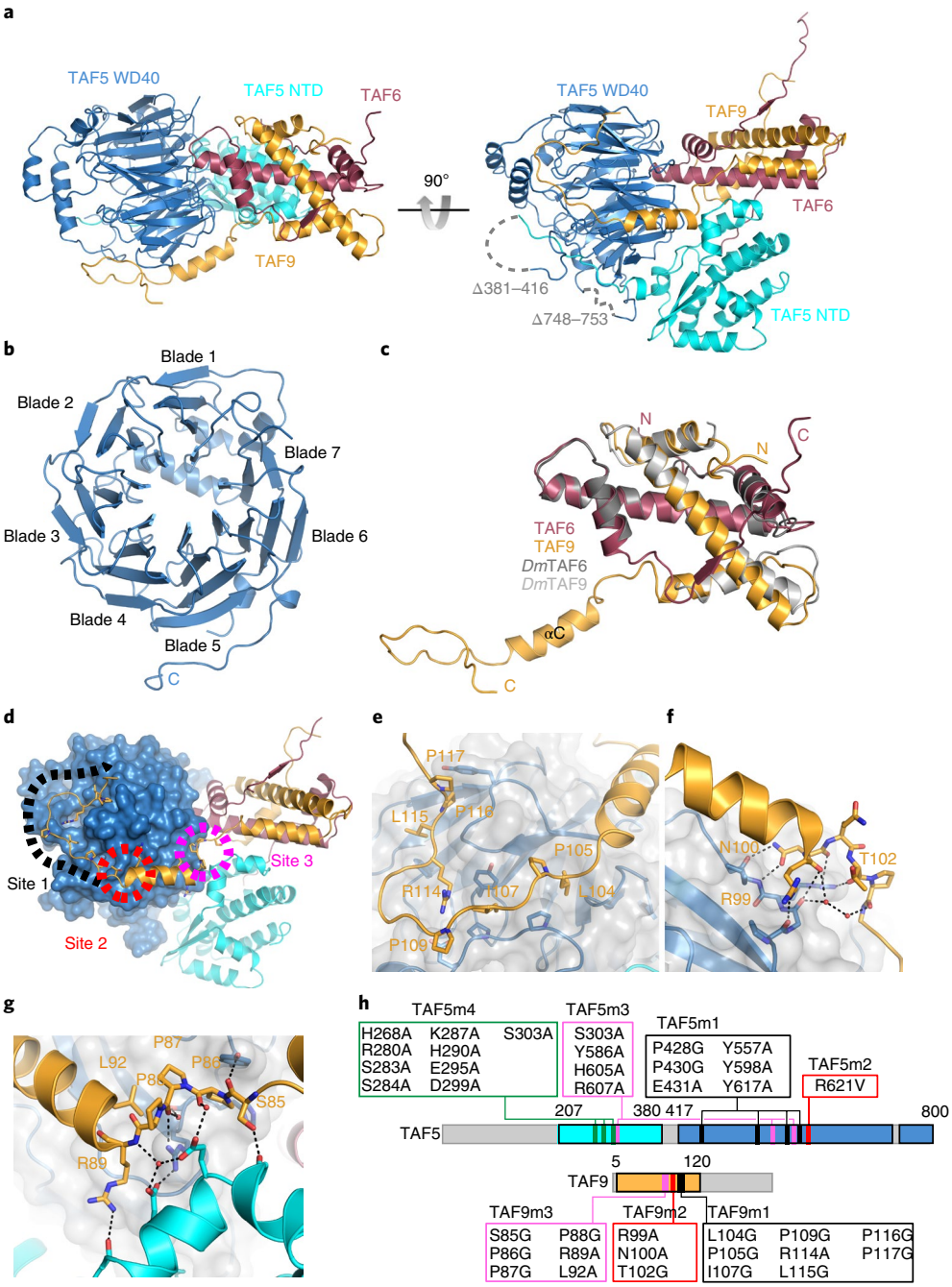

Fig. 2 | Crystal structure of the TAF5–TAF6–TAF9 complex. a, The TAF5–TAF6–TAF9 complex is shown in two orientations. The TAF6–TAF9 HFD heterodimer is intimately wedged in between the TAF5 NTD and WD40-repeat domain. The TAF5 NTD and WD40-repeat domain are colored in cyan and blue, respectively. TAF6 is colored in red, TAF9 is colored in orange. Disordered loops are drawn as a gray dashed lines. **b**, The TAF5 WD40-repeat domain is viewed from its bottom cavity, adopting a seven-bladed β-propeller. **c**, The human TAF6–TAF9 is superimposed on the TAF6–TAF9 HFD dimer from *D. melanogaster* (PDB ID 1TAF[29]). *D. melanogaster* (*Dm*) proteins are colored in grey, human TAF6 in red and human TAF9 in orange. Note the extended C-terminal domain in human TAF9. **d**, Overview of the structure with the three major interactions anchoring TAF9 to TAF5. **e–g**, Close-up showing the TAF9 C-terminal loop (**e**), hydrogen-bond network at the C-terminus of the αC helix (**f**) and triple proline turn (**g**). **h**, Mutations to investigate TAF5–TAF9 interactions. Colors as in **d**; green indicates the mutated region in the TAF5 NTD that target the TAF5–TAF6–TAF9 interface.

Supplementary Fig. 6a–d). TAF9m1 did not incorporate into TFIID, although it retained a low, albeit significant, interaction with TAF6. Whereas the incorporation was strongly reduced compared to wild-type TAF9, TAF9m1 incorporation into a (partial) SAGA complex was indicated by the significantly enriched SAGA subunits TADA3, TADA2B, SUPT7L, SUPT20H, KAT2A and TRRAP (Supplementary Fig. 6a). TAF9m2 disrupted TFIID formation to a similar extent as TAF9m1, while maintaining pairing with TAF6. This TAF9 mutant had no noticeable effect on SAGA. Similarly, TAF9m3 compromised TFIID assembly. Notably, TAF9m3 efficiently enriched most subunits of SAGA, with the exception of the subunits forming the deubiquitination module[9], suggesting a role for the region that is mutated in TAF9m3 in mediating deubiquitination integration. Finally, TAF9m4 targeting all TAF9–TAF5 interfaces abolished

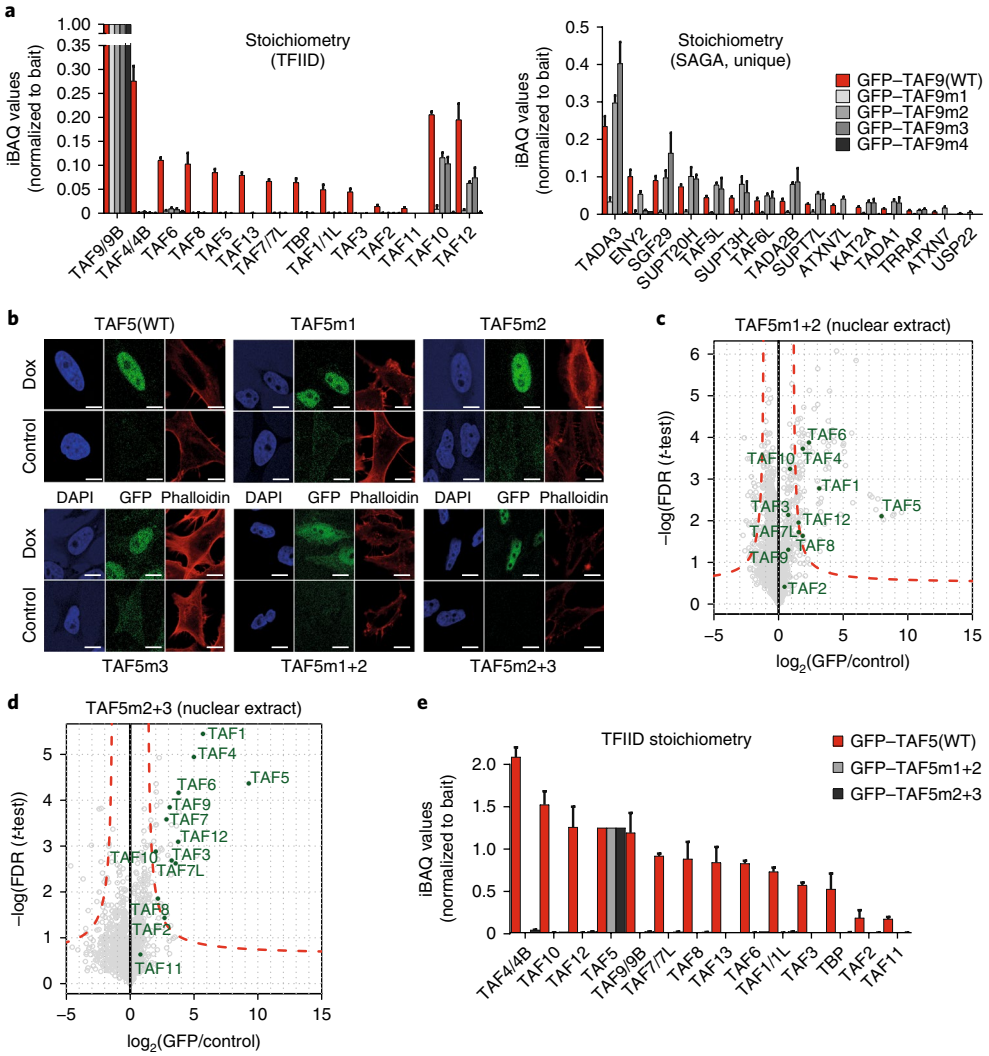

Fig. 3 | TAF5–TAF9 interactions are analysed using mutagenesis and quantitative proteomics. a, Stoichiometry plots of TFIID and SAGA subunits from GFP–TAF9 co-IPs are shown, normalized to bait protein. Data are mean ± s.d. of technical triplicates. **b**, Confocal fluorescence microscopy reveals nuclear localization of all GFP–TAF5 proteins studies except for the GFP–TAF5m1 and GFP–TAF5m2 double mutant, which is also present in the cytoplasm. Scale bars, 10 µm. **c–e**, Nuclear co-IPs of GFP–TAF5, GFP–TAF5m1 and GFP–TAFm2 (TAF5m1 + 2) and GFP–TAF5m2 and GFP–TAG5m3 (TAF5m1 + 2) mutants show that TFIID assembly is compromised when the mutant proteins are expressed. Each data point in the volcano plots is plotted as the mean of technical triplicates. Dashed red lines denote the threshold between background and significant enrichment (two-tailed Student's *t*-test; FDR = 1%; $S_0 = 1$). Data in the stoichiometry plot represent the mean ± s.d. of technical triplicates. $n = 2$ independent experimental replicates for the representative GFP–TAF5m1 and GFP–TAF52 sample. Source data for **a** and **e** are available in the online version of the paper.

both TFIID and SAGA formation completely and this mutant associated only with TAF6. In human SAGA, TAF9 pairs with the TAF6 paralogue TAF6L. Of note, TAF6L was not observed in TAF9m1 (and TAF9m4) co-IPs. Our results demonstrate that mutation of TAF9 residues critical for the TAF5–TAF9 interaction completely disrupt TFIID formation. By contrast, the TAF9 mutations affect SAGA formation to a much lower extent, indicating that the conformation adopted by TAF9 and its interactions may be substantially different within TFIID compared compared to the SAGA complex.

We introduced mutations in TAF5 reciprocal to those in the TAF9 mutants for validation (Figs. 2h, 3b–e and Supplementary Figs. 6e–g, 7). First, we included mutations in TAF5, targeting each of the three distinct interface regions individually. Notably, co-IPs and qMS of GFP-tagged versions showed that the three individual TAF5 mutants did not completely disrupt TFIID but that the relative abundance of TFIID subunits was reduced compared to wild-type TFIID, indicating that the formattion of TFIID was less efficient (Supplementary Fig. 6f). This is not because of reduced protein levels, because all TAF5 mutants were expressed at levels higher than those of wild-type TAF5 (Supplementary Fig. 6g). We next prepared TAF5 mutants, each combining two sets of mutations. The combinations of TAF5m1 and TAFm2 (P428G, P430G, Q431A, Y557A, Y598A, Y617A and R621V) and TAF5m2 and TAF5m3 (S303A, Y586A, H605A, R607A, andR621V) completely prevented TFIID formation (Fig. 3c–e).

Our mutational analyses validated the atomic interactions that were observed in our crystal structure, confirming their central importance for TFIID assembly and integrity in vivo. Moreover, our results with mutants of TAF9 suggest that there are distinct interactions of this shared subunit with TFIID and SAGA, implicating TAF9 structural dynamics in the formation pathways of the holocomplexes.

Chaperonin CCT interacts with the WD40 domain of TAF5. Cells that expressed GFP–TAF5m1 and GFP–TAF5m2 exhibited a marked GFP signal in the cytoplasm, in contrast to the predominantly

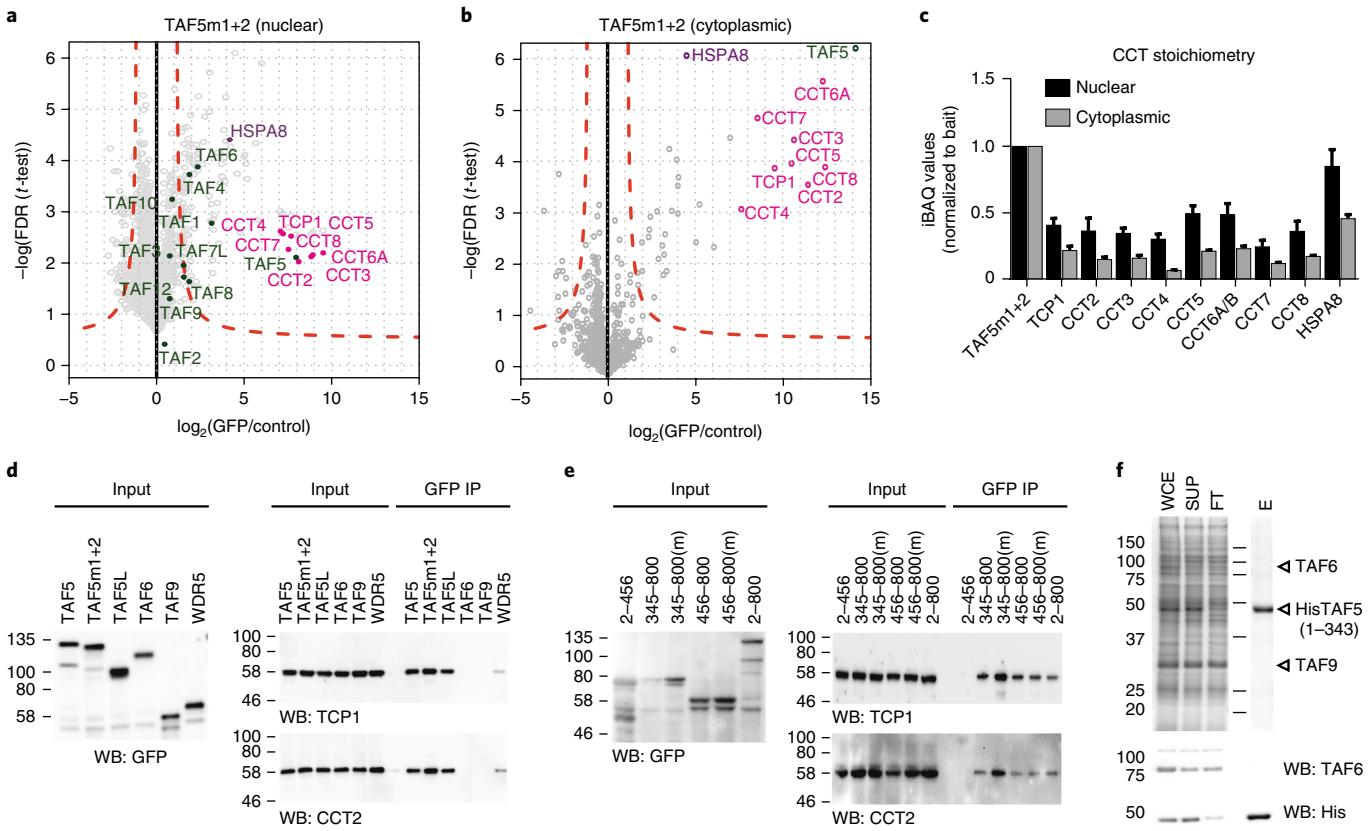

Fig. 4 | The chaperonin CCT engages TAF5. a–c, The GFP–TAF5m1 and GFP–TAF5m2 mutant enriches all subunits of CCT in almost stoichiometric ratios. Chaperonin subunits are colored in magenta, TFIID subunits in green and HSPA8 in purple. Relative stoichiometry is normalized to bait. Each data point in the volcano plots is plotted as the mean of technical triplicates. Dashed red lines denote the threshold between background and significant enrichment (two-tailed Student's *t*-test; FDR = 1%; $S_0 = 1$). Data in the stoichiometry plot represent mean ± s.d. of technical triplicates. $n = 2$ independent experimental replicates for the representative GFP–TAF5m1 and GFP–TAF52 nuclear sample. **d**, GFP–TAF interactions with CCT in whole-cell extracts from transient-transfection experiments were analysed by immunoblot. Input represents 1% of the protein sample used in the co-IPs. Hallmark CCT subunits TCP1 and CCT2 were analysed. $n = 3$ technical replicates. **e**, Transient-transfection experiments with whole-cell extract GFP fusions of wild-type (WT) TAF5 or GFP–TAF5m1 and GFP–TAF5m2 (m), TAF5(2–456) (ΔWD40), TAF5(345–800) (ΔNTD) and TAF5(456–800) (WD40) probed by co-IPs and immunoblot. $n = 2$ technical replicates. **f**, Talon pull-down of co-expressed TAF6, TAF9 and truncated TAF5 analyzed by SDS-PAGE (left) and immunoblot (right). E, eluted fraction of His for oligohistidine tag; FT, flow-through; SUP, cleared lysate; WB, western blot; WCE, whole-cell extract. Source data for **c** are available in the online version of the paper. Uncropped blot/gel images are shown in Supplementary Data 1.

nuclear localization of fluorescence of the other mutant and wild-type GFP–TAF5 cell lines (Fig. 3b). Notably, TAF5m1 and TAF5m2 co-IPs revealed eight highly enriched proteins in close to equimolar amounts, which we identified as the subunits of the CCT complex (Fig. 4a,b). CCT assists in the folding of about 10% of the proteome, including many WD40-repeat proteins[16,17,30]. In addition, strong enrichment of the HSPA8 protein was observed (Fig. 4a,b), suggesting that HSPA8 is involved in TAF5 delivery to the CCT chaperonin[31]. We reasoned that CCT may facilitate folding of the TAF5 C-terminal WD40-repeat domain to allow handover to the TAF6–TAF9 heterodimer, as an essential step in holo-TFIID formation. Our qMS experiments show that the TAF5m1 and TAF5m2 double mutant is deficient in accreting TFIID subunits including TAF6–TAF9 (Fig. 4a–c). We hypothesized that the TAF5m1 and TAF5m2 protein may be retained in the CCT folding chamber, possibly owing to impaired interactions with TAF6–TAF9. We investigated whether our observation only related to this particular mutant. Careful inspection confirmed the presence of CCT, albeit in lower amounts, in all GFP–TAF5 mutant co-IPs (Supplementary Fig. 8a,b). Notably, CCT was similarly detected in co-IPs of wild-type GFP–TAF5 from the cytoplasm, but it was absent in the nuclear extract, suggesting that CCT interactions occurred transiently with newly synthesized TAF5 (Supplementary Fig. 8b). Similar co-IPs using GFP-tagged TBP, TAF1, TAF6, TAF7 and TAF9 proteins did not contain any of the CCT subunits (data not shown). We conclude that the CCT interaction is highly specific for the TAF5 subunit of TFIID.

We used transiently transfected HEK293T cells to substantiate our findings. Overexpression of GFP–TAF5, the GFP–TAF5m1 and GFP–TAF5m2 double mutant or SAGA-specific GFP–TAF5L strongly enriched CCT in co-IPs, unlike GFP-tagged TAF6 and TAF9 or WDR5, a transcriptional regulator with a comparable WD40-repeat domain (Fig. 4d), confirming that TAF5 and its close paralogue TAF5L are bona fide CCT substrates. Note that WDR5 can be produced in *Escherichia coli*, indicating that WDR5 folding does not depend on a eukaryotic chaperone[32]. Transient overexpression of TAF5 fragments confirmed that the WD40 domain is mediating the CCT interaction and that the NTD did not interact with CCT subunits (Fig. 4e). Recombinant reconstitution also confirmed that the TAF5–TAF6–TAF9 complex is dependent on the TAF5 WD40 domain (Fig. 4f).

CCT hands over nascent TAF5 to TAF6–TAF9 for holo-TFIID assembly. Formation of the CCT–TAF5 complex would presumably occur following translation in the cytoplasm. We tested this in pulse–chase experiments in our Dox-inducible cell lines (Fig. 5a–c

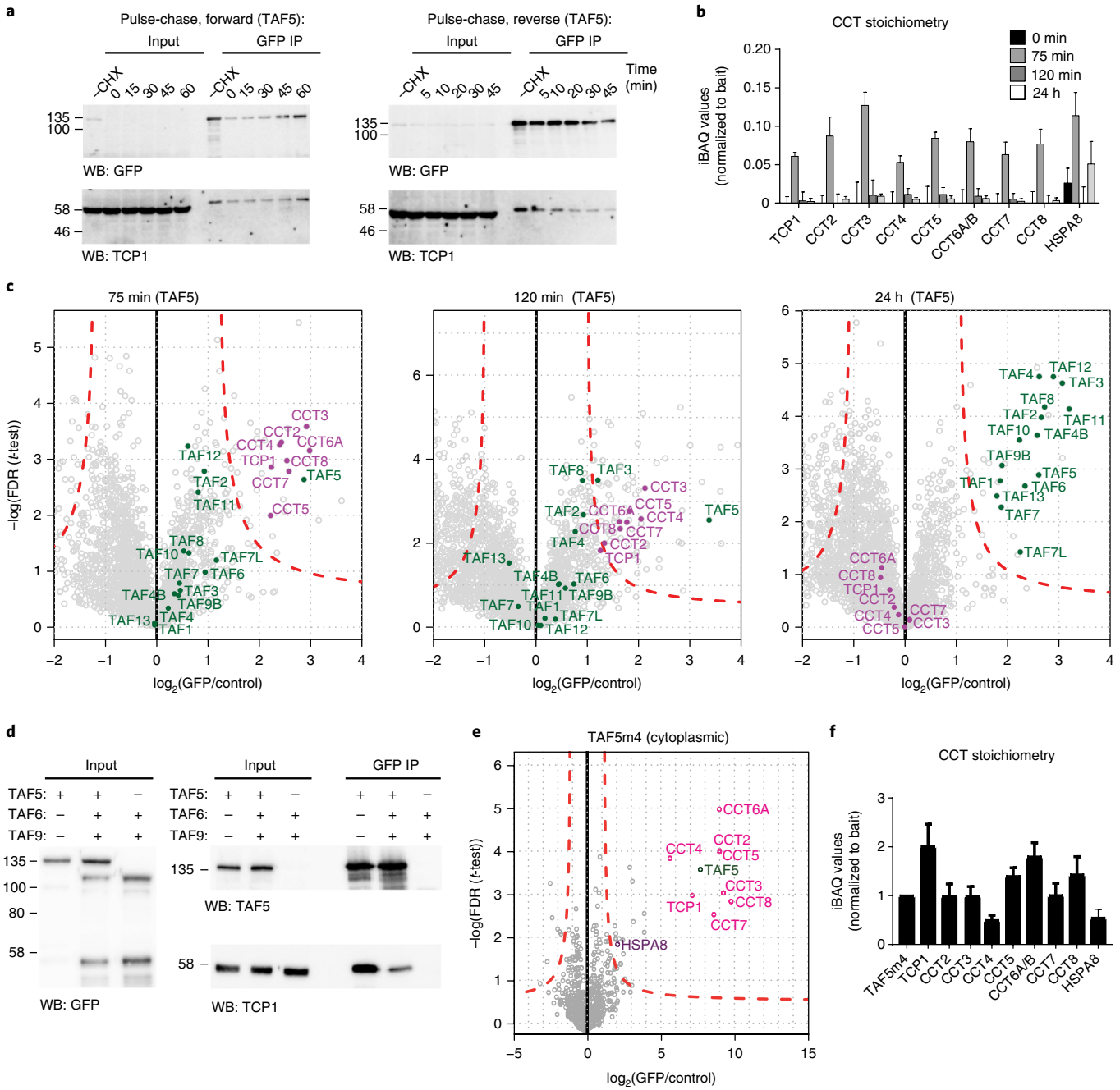

Fig. 5 | CCT is required for holo-TFIID assembly. a, CCT interaction with newly synthesized GFP–TAF5 was analysed by pulse–chase experiments. Input represents 5% of the protein sample used in co-IPs. **b**,**c**, Pulse–chase experiment sampled at different time points. CCT binding to TAF5 peaks at 75 min after inducing GFP–TAF5 expression. Relative abundance was normalized to bait. Baseline was taken at $t=0$. CCT subunits are colored in magenta and TFIID subunits in green. Each data point in the volcano plots is plotted as the mean of technical triplicates. Dashed red lines denote the threshold between background and significant enrichment (two-tailed Student's t-test; FDR = 1%; $S_0=1$). Data in the stoichiometry plot represent the mean ± s.d. of technical triplicates. **d**, Transient co-transfection experiments with GFP–TAF5, GFP–TAF6 and GFP–TAF9 analysed by co-IPs and immunoblot. TAF6–TAF9 is required for the release of TAF5 from CCT. $n=2$ technical replicates. Uncropped blot/gel images are shown in Supplementary Data 1. **e**,**f**, qMS analysis of GFP–TAF5m4 co-IPs obtained from the cytoplasmic fraction. The stoichiometry plot is normalized to bait. Each data point in the volcano plots is plotted as the mean of technical triplicates. Data in the stoichiometry plot represent the mean ± s.d. of technical triplicates. Source data for **b** and **f** are available in the online version of the paper.

and Supplementary Fig. 9a,b). Before adding Dox, cycloheximide (CHX) was applied to block translation, resulting in mRNA accumulation. In the chase phase, CHX blockade and Dox treatment were removed and nascent GFP–TAF5 was sampled over time, revealing progressive CCT binding to newly synthesized GFP–TAF5 in co-IP experiments (Fig. 5a). Experiments using the GFP–TAF5m1 and GFP–TAF5m2 double mutant yielded comparable results; GFP–TAF7 was included as a negative control (Supplementary Fig. 9b). In reverse pulse–chase experiments, prolonged Dox induction enabled GFP–TAF5 protein accumulation, followed by CHX addition to inhibit new protein synthesis. We observed time-dependent disassociation of GFP–TAF5 from CCT. By contrast, TAF5m1and

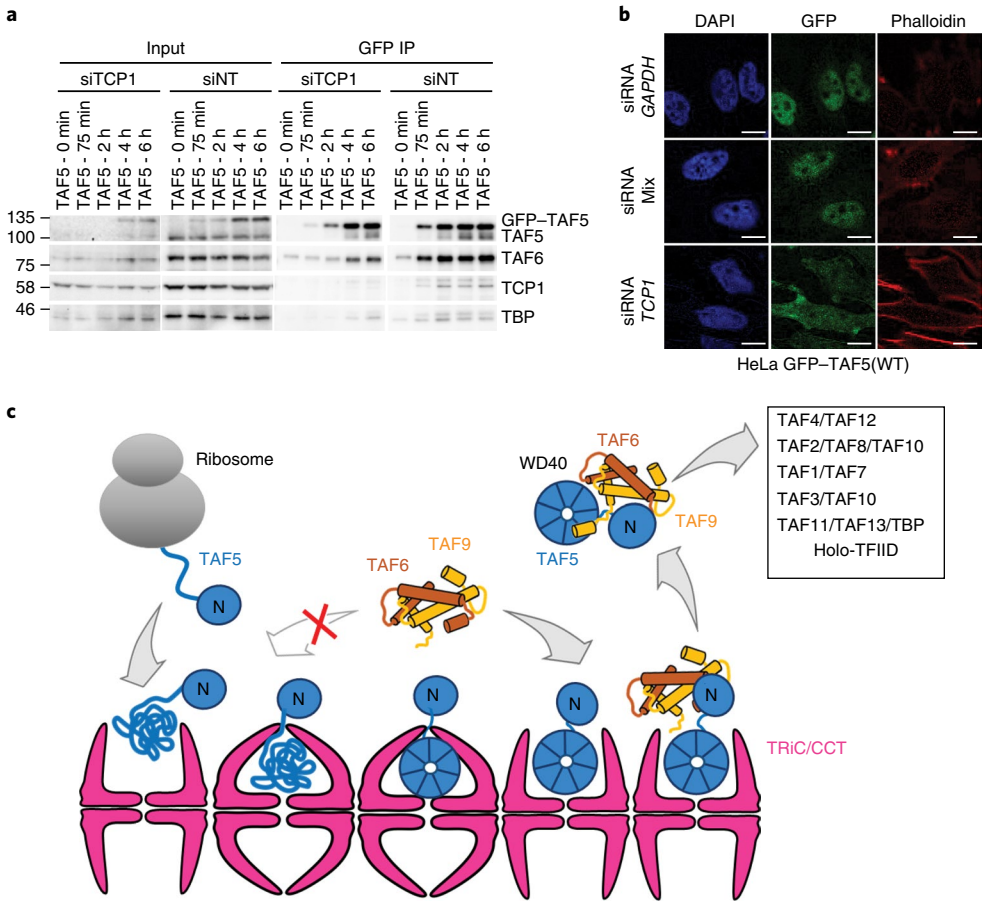

Fig. 6 | CCT chaperonin-assisted early events in TFIID assembly. a, siRNA-mediated knockdown of CCT subunit TCP1 interferes with GFP–TAF5 integration into TFIID. Control cells were treated with a non-targeting siRNA (siNT). Input represent 2% of the protein sample used in each co-IP. $n = 2$ technical replicates. Uncropped blot/gel images are shown in Supplementary Data 1. **b**, TCP1 knockdown affects nuclear localization of GFP–TAF5. Scale bars, 10 µm. $n = 2$ technical replicates. **c**, Newly synthesized TAF5 is captured by the chaperonin (left). The TAF5 N-terminal domain folds autonomously, whereas folding of the WD40-repeat domain depends on CCT. Release of readily folded TAF5 from the chaperonin requires binding of TAF6–TAF9, resulting in a discrete TAF5–TAF6–TAF9 complex in the cytoplasm. Binding of further TAFs and TBP completes the assembly of the functional holo-TFIID complex.

TAFm2 remained bound to the chaperonin. Mass spectrometry analysis of GFP–TAF5-induced cells showed that newly synthesized TAF5 enriched all CCT subunits (Fig. 5b), with CCT binding peaking at 75 min before declining, corroborating the transient nature of the interaction. Notably, co-IPs at later time points progressively enriched TAFs and TBP, whereas CCT disappeared into background, consistent with TAF5 release from CCT and incorporation into the holo-TFIID complex (Fig. 5c).

Our model indicates that the handover of TAF5 from CCT to TAF6–TAF9 would be controlled by the levels of the TAF6–TAF9 heterodimer and be dependent on the initial encounter between TAF6–TAF9 and the NTD of TAF5. Indeed, transient co-expression experiments of GFP-tagged TAF5, TAF6 and TAF9 in various combinations underscored that TAF6–TAF9 is required for TAF5 release from CCT (Fig. 5d). We designed a TAF5 mutant that had mutations in the NTD (TAF5m4) that were predicted to disrupt the interaction with TAF6–TAF9 (Fig. 2h and Supplementary Fig. 7). Subsequent co-IP of this mutant followed by qMS showed that TAF5m4 did not form TFIID submodules and that it was retained in the CCT complex (Fig. 5e,f). These data indicate that the CCT complex acts as an essential checkpoint in TFIID assembly. Next, we investigated the effect of short interfering RNA (siRNA)-mediated knockdown of *TCP1*, which affects CCT function[16]. *TCP1* knockdown in GFP–TAF5-induced cells markedly reduced co-purification of TAF6 as well as TBP, indicating that the interaction with CCT is a prerequisite to prime TAF5 for TAF6–TAF9 binding and incorporation into the TFIID complex (Fig. 6a). Consistent with this, *TCP1* knockdown affected cellular localization and a considerable proportion of GFP–TAF5 was retained in the cytoplasm (Fig. 6b).

Discussion

Taken together, we provide unique insights at the molecular level into early events in the regulated assembly of the basal transcription factor TFIID. Our results provide compelling evidence that the chaperonin CCT associates with newly synthesized TAF5 in the cytoplasm of cells, acting as an essential checkpoint for the formation of functional TFIID holo-complexes (Fig. 6c and Supplementary Video 1). Our data demonstrate that the interaction with CCT is a prerequisite for the formation of a TFIID submodule that consists of TAF5, TAF6 and TAF9. Presumably, CCT facilitates the folding of the WD40-repeat domain and stabilizes TAF5 in a conformation that is compatible with the formation of the multitude of interactions between the TAF6–TAF9 and the TAF5 NTD that were observed in the crystal structure.

Our results show that TAF6–TAF9 binding is required to release TAF5 from CCT and indicate that CCT rebinding may be prevented by docking of α-helix 4 of TAF9 on the lateral surface of the TAF5

WD40 domain (Fig. 2a and Supplementary Video 1). This handover is critical for the formation of the holo-TFIID complex that contains a full complement of TAFs and TBP. Additionally, we identify TAF5–TAF6–TAF9 as a discrete cytoplasmic TFIID subcomplex that is poised to engage other TAFs including the TAF4–TAF12 heterodimer, thus substantiating the concept of holo-TFIID assembly from preformed submodules and pinpointing a chaperonin as a key factor in this process.

Our finding that not only TAF5, but also its paralogue TAF5L interact with CCT, suggests that the handover mechanism described here is not confined to TFIID and similar mechanisms may also govern early stages of SAGA assembly. We anticipate that our observations will have general implications for the molecular mechanisms that regulate the assembly of the many multiprotein machines that are at work in the cells.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41594-018-0156-z

Received: 4 June 2018; Accepted: 25 October 2018;
Published online: 3 December 2018

References

1. Roeder, R. G. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.* **579**, 909–915 (2005).
2. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
3. Bieniossek, C. et al. The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699–702 (2013).
4. Louder, R. K. et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
5. Muller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* **20**, 533–540 (2010).
6. Helmlinger, D. & Tora, L. Sharing the SAGA. *Trends Biochem. Sci.* **42**, 850–861 (2017).
7. Baptista, T. et al. SAGA is a general cofactor for RNA polymerase II transcription. *Mol. Cell* **68**, 130–143 (2017).
8. Warfield, L. et al. Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* **68**, 118–129 (2017).
9. Spedale, G., Timmers, H. T. & Pijnappel, W. W. ATAC-king the complexity of SAGA during evolution. *Genes Dev.* **26**, 527–541 (2012).
10. Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev.* **7**, 1291–1308 (1993).
11. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
12. Ellis, R. J. Assembly chaperones: a perspective. *Phil. Trans. R. Soc. B* **368**, 20110398 (2013).
13. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).
14. Ramos, P. C. & Dohmen, R. J. PACemakers of proteasome core particle assembly. *Structure* **16**, 1296–1304 (2008).
15. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
16. Lopez, T., Dalton, K. & Frydman, J. The mechanism and function of group II chaperonins. *J. Mol. Biol.* **427**, 2919–2930 (2015).
17. Pines, A. et al. TRiC controls transcription resumption after UV damage by regulating Cockayne syndrome protein A. *Nat. Commun.* **9**, 1040 (2018).
18. Muñoz, I. G. et al. Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18**, 14–19 (2011).
19. Altelaar, A. F., Munoz, J. & Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14**, 35–48 (2013).
20. Ahrens, C. H., Brunner, E., Qeli, E., Basler, K. & Aebersold, R. Generating and navigating proteome maps using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **11**, 789–801 (2010).
21. Wright, K. J., Marr, M. T. II & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl Acad. Sci. USA* **103**, 12347–12352 (2006).
22. Trowitzsch, S. et al. Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6**, 6011 (2015).
23. Gupta, K. et al. Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *eLife* **6**, e30395 (2017).
24. Scheer, E., Delbac, F., Tora, L., Moras, D. & Romier, C. TFIID TAF6–TAF9 complex formation involves the HEAT repeat-containing C-terminal domain of TAF6 and is modulated by TAF5 protein. *J. Biol. Chem.* **287**, 27580–27592 (2012).
25. Gangloff, Y. G., Romier, C., Thuault, S., Werten, S. & Davidson, I. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* **26**, 250–257 (2001).
26. van Nuland, R. et al. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol. Cell. Biol.* **33**, 2067–2077 (2013).
27. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
28. Bhattacharya, S., Takada, S. & Jacobson, R. H. Structural analysis and dimerization potential of the human TAF5 subunit of TFIID. *Proc. Natl Acad. Sci. USA* **104**, 1189–1194 (2007).
29. Xie, X. et al. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316–322 (1996).
30. Miyata, Y., Shibata, T., Aoshima, M., Tsubata, T. & Nishida, E. The molecular chaperone TRiC/CCT binds to the Trp-Asp 40 (WD40) repeat protein WDR68 and promotes its folding, protein kinase DYRK1A binding, and nuclear accumulation. *J. Biol. Chem.* **289**, 33320–33332 (2014).
31. Cuéllar, J. et al. The structure of CCT–Hsc70$_{NBD}$ suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nat. Struct. Mol. Biol.* **15**, 858–864 (2008).
32. Han, Z. et al. Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol. Cell* **22**, 137–144 (2006).

Acknowledgements

We thank S. Trowitzsch (Goethe University Frankfurt), P. Legrand and A. Thompson (Synchrotron SOLEIL), W. Vonk (Princess Maxima Centre Utrecht), R. Baas (Netherlands Cancer Institute Amsterdam) and M. Vermeulen (Radboud University Nijmegen) for assistance and reagents. We greatly appreciate discussion with R. Sawakar (MPI for Immunobiology and Epigenetics). This research was supported by the Netherlands Organization for Scientific Research (NWO) grants 022.004.019 (S.V.A.), ALW820.02.013 (H.T.M.T.) and 184.032.201 Proteins@Work (E.C., T.Y.L., H.R.V. and A.J.R.H.), a Kékulé fellowship from the Fonds der Chemischen Industrie (M.H.), a European Research Council Advanced grant ERC-2013-340551 (L.T.), Agence Nationale de Recherche research grants ANR-10-IDEX-0002-02 and ANR-10-LABX-0030-INRT (L.T.) and a Wellcome Trust Senior Investigator Award 106115/Z/14/Z (I.B.). This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INBS-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). This research received support from BrisSynBio, a BBSRC/EPSRC Research Centre for synthetic biology at the University of Bristol (BB/L01386X/1).

Author contributions

H.T.M.T. and I.B. conceived the study with input from S.V.A., M.H., L.T. and A.J.R.H. S.V.A. carried out all cell biology experiments, assisted by M.M. and E.S. The majority of the quantitative proteomics analyses were carried out by E.C. and T.Y.L., assisted by S.V.A. R.M.v.E. and H.R.V. carried out proteomics analyses of the pulse–chase experiment. M.H. carried out all recombinant protein work, crystallization and structural analysis, assisted by L.S., K.G. and with input from I.B. S.V.A., M.H., L.T., A.J.R.H., H.T.M.T. and I.B. designed experiments, interpreted data and wrote the manuscript together with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-018-0156-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.T.M.T. or I.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

TAF5–TAF6–TAF9 complex production. The TAF5–TAF6–TAF9 full-length complex was cloned, expressed and purified using a polyprotein strategy as described previously[3]. For all other TAF5–TAF6–TAF9 complexes, coding sequences of individual subunits and truncation mutants were cloned into acceptor (pFL) and donor (pIDC and pIDK) vectors of the MultiBac system by sequence- and ligation-independent cloning and fused by in vitro Cre–*loxP* recombination to yield a single plasmid with multiple expression cassettes[33]. Constructs used are listed in Supplementary Table 1. A decahistidine tag followed by a tobacco etch virus (TEV) NIa protease cleavage site was placed at the N terminus of the TAF5 subunit. Recombinant baculovirus was produced as previously described[34] and used to infect SF21 insect cells (Invitrogen) at a cell density of 1.0×10^6 per ml in SF4 medium. Cells were collected 72–96 h after proliferation arrest by centrifugation at 4,000*g* for 15 min. Cell pellets were resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM imidazole, 1 mM β-mercaptoethanol and 0.1 % NP-40), incubated on ice for 30 min and the lysate was cleared by centrifugation at 25,000 r.p.m. in a JA-25.50 rotor (Beckman) for 90 min. The supernatant was mixed with Talon resin (Clontech), equilibrated in Talon A buffer (25 mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM imidazole) and incubated for 1 h at 4 °C. The resin was washed with 15 column volumes of Talon A buffer, 10 column volumes of Talon HS buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl and 7.5 mM imidazole), and 15 column volumes of Talon A buffer before eluting the bound protein complex with 5 column volumes of Talon B buffer (25 mM Tris-HCl pH 7.5, 200 mM NaCl and 300 mM imidazole). Then, 6×His-TEV protease (prepared in-house) was added at a 1:100 w/w ratio and the protein was dialyzed against a 100-fold excess of dialysis buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl and 5 mM β-mercaptoethanol). The dialyzed protein was passed over a Ni-NTA resin (QIAGEN) equilibrated in dialysis buffer and the flow-through was collected. The flow-through was then passed through a MonoQ 5/50GL column (GE Healthcare). The protein complex was found in the flow-through and separated on a Superdex S200 16/60 column (GE Healthcare) equilibrated in SEC buffer (10 mM HEPES-NaOH pH 7.5, 150 mM NaCl and 1 mM dithiothreitol (DTT)). Peak fractions were pooled and concentrated to 10 mg ml^{-1}. Small aliquots were frozen in liquid nitrogen and stored at −80 °C until further use. In TAF5ΔWD40–TAF6–TAF9 pull-down experiments with full-length TAF6 and TAF9 co-expressed with a TAF5 truncation (TAF5(1–343)) mutant that lacked the WD40 domain (Fig. 4e), samples were separated by SDS–PAGE, blotted and probed using specific antibodies (L.T. laboratory), according to standard immunoblot procedures.

Crystallization, data collection and structure determination. The TAF5(194–800)–TAF6(1–92RLRRRAH)–TAF9(1–120) complex was crystallized in 0.1 M Tris-HCl pH 6.8, 0.8 M sodium citrate and 0–0.15 M NaCl using the hanging drop vapor diffusion technique by mixing 2 µl of protein solution at 5.4–7.5 mg ml^{-1} with 1 µl of reservoir solution and equilibrated against 500 µl of reservoir solution at 4 °C. Optionally, crystallization drops were seeded with 0.5 µl of a 1:10,000 to 1:1,000,000-diluted micro-seed stock solution, which was prepared by crushing a single crystal in 10 µl reservoir solution. Crystals of pyramidal shape first appeared after four days and reached their maximal size of 300 × 300 × 300 µm after ten days.

For data collection at cryogenic temperatures, crystals were transferred sequentially into 2-µl drops containing 0.1 M Tris-HCl pH 6.8, 0.15 M NaCl and increasing concentrations of sodium citrate from 1.0 M to 1.6 M before flash-freezing in liquid nitrogen. For experimental phasing, crystals of the complex were transferred to a 2-µl drop containing a stabilizing solution of 0.1 M Tris-HCl pH 6.8, 1.0 M sodium citrate and 0.15 M NaCl. Solid Ta_6Br_{12} clusters were directly added to the crystallization drop containing the pre-equilibrated crystals. After 16 h, the Ta_6Br_{12}-soaked crystals were sequentially transferred into 2-µl drops containing 0.1 M Tris-HCl pH 6.8 with 0.15 M NaCl and increasing concentrations of sodium citrate from 1.2 M to 1.6 M before flash-freezing in liquid nitrogen. Native datasets were collected at beamline ID14-4 (ESRF) at 100 K (1.0044 Å wavelength). Datasets of Ta_6Br_{12}-soaked TAF5–TAF6–TAF9-complex crystals were collected at beamline ProximaI (SOLEIL) following an optimized inverse-beam single-wavelength anomalous diffraction (SAD) data-collection strategy at the peak wavelength for Ta (1.25439 Å). The Ta_6Br_{12}-soaked crystals were non-isomorphous to native TAF5–TAF6–TAF9-complex crystals.

All diffraction data were processed with the XDS software package[35,36]. The structure of the TAF5–TAF6–TAF9 complex was solved using the Ta_6Br_{12} SAD dataset. Three Ta_6Br_{12} cluster sites were identified by HySS as implemented in PHENIX software suite[37]. The heavy-atom substructure was further refined and phases were calculated with PHASER[38]. The initial electron density map improved substantially by density modification in PARROT[39], allowing placement of four truncated molecules of the human TAF5-NTD domain (Protein Data Bank (PDB) ID 2NXP) and four molecules of the *D. melanogaster* TAF6–TAF9 HFD pair (PDB ID 1TAF) manually into the electron density map. Subsequently, the experimental electron density map was further improved by non-crystallographic symmetry averaging in RESOLVE[40,41]. The improved map allowed placement of four molecules of a seven-bladed WD40-repeat model obtained from the PHYRE server[42] (based on the TAF5 protein sequence) by molecular replacement in MOLREP[43]. This initial model was used to phase the native dataset by molecular replacement in MOLREP. After automated model building in ARP/wARP[44] and BUCCANEER[39], the model was manually adjusted using repetitive rounds of refinement in PHENIX and model building in COOT. Translation/Libration/Screw refinement was used in the final rounds of refinement with 61 individual Translation/Libration/Screw groups as determined by PHENIX.

Pull-down assay of TAF5(NTD)–TAF6–TAF9. Cell pellets were resuspended in buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM imidazole and EDTA-free complete protease inhibitor (Roche)), lysed by two cycles of freeze–thawing in liquid nitrogen and the lysate was cleared by centrifugation at 19,000*g* in a Thermo Scientific Fiberlite F21-8 x 50y fixed-angle rotor for 90 min. The supernatant was mixed with Talon resin (Clontech), equilibrated in buffer A (25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM imidazole) and incubated for 1 h at 4 °C. The resin was washed with 15 column volumes of buffer A, followed by 10 column volumes of buffer HS (25 mM Tris-HCl pH 8.0, 1,000 mM NaCl and 5 mM imidazole) and 15 column volumes of buffer A before eluting the bound protein complex with 5 column volumes of buffer B (25 mM Tris-HCl pH 8.0, 150 mM NaCl and 200 mM imidazole). Different samples were analyzed by SDS–PAGE analysis and western blot analysis using an HRP-conjugated anti-His antibody (Sigma-Aldrich) for TAF5(NTD) and an anti-TAF6 antibody (primary, from L.T.) followed by HRP-conjugated anti-mouse antibody (secondary, Sigma-Aldrich) for TAF6.

GFP–TAF cell line generation. cDNA of human proteins was amplified by PCR with gene-specific primers fused to attB recombination sequences for GATEWAY cloning. The amplified sequence was recombined into the pDON201 donor vector according to the manufacturer's protocol (ThermoFisher). The cloned sequence was verified with Sanger sequencing and was transferred to the pcDNA5-FRT-TO-N-GFP[26] Gateway destination vector by LR recombination according to the manufacturer's protocol (ThermoFisher). TAF9- and TAF5-mutant cDNAs were purchased from GenScript. Both TAF5m1 and TAF5m2 and TAF5m2 and TAF5m3 were obtained in a single round of mutagenesis PCR. All destination vectors were co-transfected with a pOG44 plasmid that encodes the Flp recombinase into HeLa Flp-In/T-REx cells[26] using polyethyleneimine transfection to generate stable Dox-inducible expression cell lines.

Cell culture methods. HeLa Flp-In/T-REx cells, which contained the Flp recombination target site and expressed the Tet repressor, were grown in Dulbecco's modified Eagle's medium (DMEM), 4.5 g l^{-1} glucose, supplemented with 10% v/v fetal bovine serum, 10 mM L-glutamine and 100 U ml^{-1} penicillin–streptomycin (all purchased from Lonza), together with 5 µg ml^{-1} blasticidin S (InvivoGen) and 200 µg ml^{-1} zeocin (Invitrogen), which were used to select for the FRT and Tet repressor, respectively. All cell lines were mycoplasma negative.

Recombined cells were selected by replacing zeocin with 250 µg ml^{-1} hygromycin B (Roche Diagnostics) 48 h after polyethyleneimine transfection. Expression of GFP-tagged protein was induced by addition of 1 µg ml^{-1} Dox, for 16–18 h. HEK293T cells used for transient-expression experiments were grown in DMEM, 4.5 g l^{-1} glucose, supplemented with 10% v/v fetal bovine serum, 10 mM L-glutamine and 100 U ml^{-1} penicillin–streptomycin (all purchased from Lonza).

Immunoblotting procedures. Cell were seeded in six-well dishes at 30,000 cells per well and induced with Dox for 16 h to 18 h before collection. Cell lysates were prepared in 1× sample buffer (160 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol and 0.05% bromophenol blue). Equal amounts of cell lysates were separated by a 10% SDS–PAGE and transferred onto a PVDF membrane. GFP purification samples for immunoblotting were collected by elution with sample buffer. The membrane was developed with the appropriate antibodies and ECL. Immunoblots were analyzed using a ChemiDoc imaging system (BioRad). The images were subjected to linear contrast/brightness enhancement in Photoshop (CS6, 13.0.6 ×64, extended) when needed for data presentation purposes. Antibodies used in the study include: GFP (JL-8, Clontech), TCP1 (91A, ThermoFisher), CCT2 (kind gift from W. Vonk/J. Frydman), GAPDH (clone 6C5, mAb374, Millipore), Penta His (QIAGEN), TAF5 (in-house, 25TA 2G7) and TAF6 (in-house, 1TA-1C2).

Extract preparation and GFP co-IP. Cells were seeded in 15-cm dishes (Greiner Cellstar, Sigma-Aldrich) and grown to 70–80% confluence prior to Dox induction. GFP–protein expression was verified using EVOS fluorescence microscopy (Thermo Fischer). Next, induced cells were collected and nuclear and cytoplasmic extracts were obtained using a modified version of the Dignam and Roeder procedure[45]. Protein concentrations were determined by Bradford assay (BioRad). Then, 1 mg of nuclear or 3 mg of cytoplasmic extract was used for GFP-affinity purification as previously described[46]. In brief, protein lysates were incubated in binding buffer (20 mM HEPES-KOH pH 7.9, 300 mM NaCl, 20% glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40, 0.5 mM DTT and 1× Roche protease inhibitor cocktail) on a rotating wheel for 1 h at 4 °C in triplicates with GBP-coated agarose beads (Chromotek) or control agarose beads (Chromotek). The beads were washed twice with binding buffer containing 0.5% NP-40, twice with PBS containing 0.5% NP-40 and twice with PBS. On-bead digestion of bound proteins was performed overnight in elution buffer (100 mM Tris-HCl pH 7.5, 2 M urea and 10 mM DTT) with 0.1 µg ml^{-1} of trypsin at room temperature and eluted tryptic peptides were bound to C18 stagetips prior to mass spectrometry analysis.

Whole-cell extracts were collected in lysis buffer (50 mM Tris-HCl pH 7.9, 10% glycerol, 100 mM NaCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1% NP-40 and 1× complete protease inhibitor cocktail (Roche)), by incubating the cells with the lysis buffer on ice for 10 min. Next, the protein samples were collected by manual harvesting and centrifuged at 14,000 r.p.m. for 20 min at 4 °C. Supernatants were snap-frozen and stored at −80 °C.

Mass spectrometry and data analysis. Tryptic peptides were eluted from the C18 stagetips in H_2O:acetonitril (50:50) with 0.1% formic acid and dried prior to resuspension in 10% formic acid. A third of this elution was injected into a Q-Exactive (Thermo Fischer) in tandem mass spectrometry mode with 90 min total analysis time. Blank samples consisting of 10% formic acid were run for 45 min between GFP and non-GFP samples, to avoid carry-over between runs. The raw data files were analyzed with MaxQuant software (version 1.5.3.30) using the Uniprot human FASTA database[46,47]. Label-free quantification values and match between run options were selected. The iBAQ algorithm was also activated for subsequent relative protein abundance estimations[27]. The obtained protein files were analyzed by Perseus software (MQ package, version 1.5.4.0), in which contaminants and reverse hits were filtered out. Protein identification based on non-unique peptides as well as proteins identified by only one peptide in the different triplicates were excluded to increase protein prediction accuracy.

For identification of the bait interactors, label-free quantification intensity-based values were transformed on the logarithmic scale ($\log_2$) to generate Gaussian distribution of the data. This enables the imputation of missing values based on the normal distribution of the overall data (in Perseus, width = 0.3; shift = 1.8). The normalized label-free quantification intensities were compared between grouped GFP triplicates and non-GFP triplicates, using a 1% permutation-based false discovery rate (FDR) in a two-tailed Student's *t*-test. The threshold for significance (S_0), based on the FDR and the ratio between GFP and non-GFP samples, was kept at the constant value of 1 for comparison purposes. Relative abundance plots were obtained by comparison of the iBAQ values of GFP interactors. The values of the non-GFP iBAQ values were subtracted from the corresponding proteins in the GFP pull-down and were next normalized to a chosen co-purifying protein for scaling and data-presentation purposes[27].

siRNA-mediated knockdown experiments. Cells were transfected using HiPerfect and a reverse-transfection protocol that was based on the manufacturer's instructions (QIAGEN). After 72 h, cells were collected for immunoblotting, mRNA analysis or immunofluorescence confocal microscopy. Protein lysates were collected in sample buffer. RNA was extracted using the RNeasy kit following the manufacturer's protocol including a DNase treatment step (QIAGEN). For immunofluorescence analysis, cells grown on coverslips were fixed in 4% paraformaldehyde in PBS for 20 min at room temperature and stored in 0.4% paraformaldehyde in PBS at 4 °C before analysis.

RT–qPCR analysis. cDNA for RT–qPCR analysis was synthesized from 500 ng of the total RNA using Superscript II (Invitrogen) with the supplied random primers mix. RT–qPCR was performed in a 25-µl final volume using iQ SYBR Green Supermix (Bio-Rad). qPCR analysis was performed in CFX Connect Real-Time PCR Detection System (Bio-Rad) and extracted data were analyzed in Excel following the $\Delta\Delta C_q$ method.

Confocal microscopy. Paraformaldehyde-fixed cells were subjected to a short cell membrane permeabilization step with 0.5% Triton X-100 in PBS for 5 min. After quenching the cross-linking with 50 mM glycine, the samples were blocked for 30 min at room temperature with 5% natural goat serum and subsequently incubated with primary antibodies for 2 h at room temperature. After three sequential washing steps with PBS, secondary antibodies conjugated to a fluorophore were added, together phalloidin (Abcam) for staining of actin for 45 min at room temperature. For nuclear staining, the samples were incubated with DAPI (2 mg l^{-1}, Sigma-Aldrich) for 5 min before mounting the coverslips on microscopy slides (ThermoFischer Scientific) with Immu-Mount (Invitrogen).

Images were obtained using a SP8-X confocal microscope (Leica Microsystems) using a HC PL APO 63×/1.40 NA oil CS2 objective. Gain and offset settings were adjusted according to the fluorescence signal, but they were kept constant in comparative experimental designs such as Dox-induction tests. Exported files were next subjected to linear contrast and brightness processing in Photoshop (CS6, 13.0.6 ×64, extended) for image presentation purposes.

Pulse–chase experiments. *Forward setup (tracing newly synthesized proteins).* Cells were seeded in 10-cm culture dishes and grown to approximately 90% confluence. Next, protein translation was blocked by addition of 50 µg ml^{-1} CHX (Sigma-Aldrich). Then, 30 min after CHX block, cells were induced with 1 µg ml^{-1} Dox for 6 h to accumulate GFP–protein mRNA. Prior to sample collection, cells were washed with PBS twice for complete removal of CHX and medium without Dox was refreshed for the duration of the chosen time points. Whole-cell extracts were collected for each of the time points and used for GFP affinity purification. Enriched proteins were separated by SDS–PAGE and detected by immunoblotting.

Reverse setup (blocking new protein synthesis). Cells were grown in 10-cm culture dishes to 70% confluence before Dox induction. Then, 50 µg ml^{-1} of CHX was added after 16 h of induction to block the synthesis of new proteins and to allow gradual disassociation of the newly synthesized proteins from their respective folding partners. Cells were collected at different time points after translational block induction and used for GFP affinity purification in order to analyze associated proteins. Samples were analyzed by SDS–PAGE and immunoblotting.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and crystallographic structure factors have been deposited in the PDB under accession code 6F3T. Proteomics data have been deposited in the PRIDE database under accession code PXD011293. Source data for Figs. 1, 3–5 and Supplementary Figs. 6, 8 and 9 are available in the online version of the paper. All other data supporting findings in this study are available from the corresponding authors upon reasonable request.

References

33. Haffke, M., Viola, C., Nie, Y. & Berger, I. Tandem recombineering by SLIC cloning and Cre–LoxP fusion to generate multigene expression constructs for protein complex research. *Methods Mol. Biol.* **1073**, 131–140 (2013).
34. Fitzgerald, D. J. et al. Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021–1032 (2006).
35. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
36. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
37. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
38. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
39. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
40. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
41. Terwilliger, T. C. Automated structure solution, density modification and model building. *Acta Crystallogr. D* **58**, 1937–1940 (2002).
42. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* **4**, 363–371 (2009).
43. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
44. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* **3**, 1171–1179 (2008).
45. Carey, M. F., Peterson, C. L. & Smale, S. T. Dignam and Roeder nuclear extract preparation. *Cold Spring Harbor Protoc.* **2009**, pdb.prot5330 (2009).
46. Spruijt, C. G., Baymaz, H. I. & Vermeulen, M. Identifying specific protein–DNA interactions using SILAC-based quantitative proteomics. *Methods Mol. Biol.* **977**, 137–157 (2013).
47. Low, T. Y. et al. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Rep.* **5**, 1469–1478 (2013).

nature research

Corresponding author(s): Berger

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Native X-ray datasets were collected at beamline ID14-4 (ESRF, Grenoble, France), datasets of Ta6Br12 soaked TAF5/TAF6/TAF9 complex crystals were collected at beamline Proximal (Soleil, Gif-sur-Yvette, France) using their associated publicly available software packages. All proteomics data were collected at common commercial mass spectrometers utilitizing their publicly available associated software packages.

Data analysis	All X-ray diffraction data were processed with the XDS software package, Ta6Br12 cluster sites were identified by HySS as implemented in PHENIX software suite (Version 1.12_2829), phases were calculated with PHASER (PHENIX software suite Version 1.12_2829), density modification was carried out in PARROT (Version 1.0.2), NCS averaging in RESOLVE (PHENIX software suite Version 1.12_2829), WD40 repeat model was obtained from the PHYRE server, molecular replacement was performed in MolRep (Version 11.0.05), automated model building was made in ARP/wARP (Version 7.4) and BUCCANEER (Version 1.5), model building was carried out in COOT (Version 0.8.2). All software used is publicly available. For proteomics experiments, the raw data files were analyzed with MaxQuant software (version 1.5.3.30) using Uniprot human FASTA database. Intensity based absolute quantification (iBAQ) algorithm was also activated for subsequent relative protein abundance estimation. The obtained protein files were analyzed by Perseus software (MQ package, version 1.5.4.0). Real-Time PCR was carried out using the RT-PCR Detection System (Bio-Rad, Hercules, CA) and extracted data was analyzed in software Microsoft Excel. Exported files from confocal microscopy, obtained using SP8-X confocal microscope (Leica microsystems, Germany) were subjected to linear contrast and brightness processing in Photoshop (CS6, 13.0.6 x64, extended) for image representation purposes.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

X-ray structure coordinates and structure factors have been deposited in the Protein Data Bank under accession code 6F3T. All mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under the dataset identifier PXD011293.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The raw data files were analyzed with MaxQuant software (version 1.5.3.30) using label-free quantification (LFQ), match between runs (between triplicates), and the iBAQ algorithm for subsequent relative protein abundance estimation. The obtained protein files were analyzed by Perseus software (MQ package, version 1.5.4.0). For identification of the bait interactors LFQ intensity-based values were transformed on the logarithmic scale (log2) to generate Gaussian distribution of the data. This allows for imputation of missing values based on the normal distribution of the overall data (in Perseus, width = 0.3; shift = 1.8). The normalized LFQ intensities were compared between grouped GFP triplicates and non-GFP triplicates, using 1% permutation-based false discovery rate (FDR) in a two-tailed t-test. The threshold for significance (S0), based on the FDR and the ratio between GFP and non-GFP, samples was kept at the constant value of 1 for comparison purposes. Relative abundance plots were obtained by comparison of the iBAQ values of GFP interactors. The values of the non-GFP iBAQ values were subtracted from the corresponding proteins in the GFP pull-down and were next normalized on a chosen co-purifying protein for scaling and data representation purposes.
Data exclusions	Contaminants, reverse hits, non-unique peptides or only one-peptide-identified proteins were excluded from Mass Spectrometry data prior plotting.
Replication	All co-IPs for MS are performed in technical triplicates and key experiments performed in biological duplicates. Biological replicates of experiments analyzed with WB were performed at least twice.
Randomization	not applicable
Blinding	not applicable

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: All materials used int his study are readily available per common material transfer agreements (MTAs) from the authors.

Antibodies

Antibodies used: Pentahis antibody (mouse IgG1 antibody isotype, monoclonal, HHHHH epitope) from Qiagen, Hilden, Germany.
anti-GFP mouse antibody monoclonal (JL-8, Clontech)
anti-TCP1 rat monoclonal antibody (91A, ThermoFisher)
anti-hTAF6 (in-house 25TA 2G7) mouse monoclonal antibody
anti-hTAF5 (in house 1TA-1C2) mouse monoclonal antibody

Validation: Pentahis antibody was validated by utilizing a large set (>100) of recombinant proteins containing oligohistidine tags in the berger laboratory.
anti-GFP was validated validated using dox-inducible GFP-fusion of the proteins of interest.
anti-TCP1 was validated using dox-inducible GFP-fusion of the proteins of interest as well as siRNA knock-down.
anti-hTAF6 25TA 2G7 mAb has been described and validated in Bell et al. (2001) Molecular Cell, 8, 590-600.
anti-hTAF5 1TA 1C2 antibody has been described and validated in Dubrovskaya et al. (1996) EMBO J. 15, 3702-3712.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): Sf21 insect cells (Life Technologies, Carlsbad, CA, USA) for protein expression.
HeLa Flp-In/T-REx cells for generation of stable inducible GFP-fusion cell lines (see van Nuland et al, 2013)

Authentication: SF21 cell lines were authenticated by genome sequencing (see Kohler et al, Nat Methods 2016)

Mycoplasma contamination: HeLa Flp-In/T-REx-derived cell lines - negative for mycoplasma

Commonly misidentified lines (See ICLAC register): not applicable