



OPEN

# Structure of the human SAGA coactivator complex

Dominik A. Herbst<sup>1,2,8</sup>, Meagan N. Esbin<sup>3,4,8</sup>, Robert K. Louder<sup>1,4,6</sup>, Claire Dugast-Darzacq<sup>1,3</sup>, Gina M. Dailey<sup>1,3</sup>, Qianglin Fang<sup>1,2,7</sup>, Xavier Darzacq<sup>1,3</sup>, Robert Tjian<sup>3,5</sup> and Eva Nogales<sup>1,2,3,5</sup>✉

**The SAGA complex is a regulatory hub involved in gene regulation, chromatin modification, DNA damage repair and signaling. While structures of yeast SAGA (ySAGA) have been reported, there are noteworthy functional and compositional differences for this complex in metazoans. Here we present the cryogenic-electron microscopy (cryo-EM) structure of human SAGA (hSAGA) and show how the arrangement of distinct structural elements results in a globally divergent organization from that of yeast, with a different interface tethering the core module to the TRRAP subunit, resulting in a dramatically altered geometry of functional elements and with the integration of a metazoan-specific splicing module. Our hSAGA structure reveals the presence of an inositol hexakisphosphate (InsP<sub>6</sub>) binding site in TRRAP and an unusual property of its pseudo-(Ψ)PIKK. Finally, we map human disease mutations, thus providing the needed framework for structure-guided drug design of this important therapeutic target for human developmental diseases and cancer.**

Transcription of protein coding genes depends on the essential coactivators TFIID and SAGA (Spt-Ada-Gcn5 acetyltransferase)<sup>1,2</sup>. SAGA regulates gene expression by interacting with enhancer-bound activators, recruiting the transcriptional machinery and modifying promoter-proximal chromatin<sup>1</sup> and is known to be involved also in DNA damage repair and signaling<sup>3</sup>. Previous studies focused primarily on yeast SAGA (ySAGA), and the first structures of the 19-subunit ySAGA were proposed to be also representative of human SAGA (hSAGA)<sup>4,5</sup>. However, hSAGA has noticeable functional and compositional differences, indicative of a divergent architecture, that integrates metazoan-specific U2 splicing subunits and lacks the essential subunit for TATA-box binding protein (TBP) binding in yeast<sup>6</sup>. Human SAGA is a 20-subunit, 1.4-MDa complex with five functional modules (Fig. 1a): a scaffolding core that includes TBP-associated factors (TAFs); a TRRAP (Transformation/Transcription domain Associated Protein) containing a phosphoinositide-3-kinase (PI3K)-related pseudoprotein kinase (ΨPIKK); a histone acetyltransferase (HAT); a deubiquitinase (DUB) and a metazoan-specific splicing (SPL) module<sup>7</sup>. The recent structural characterization of the 19-subunit *Saccharomyces cerevisiae* and *Komagataella phaffii* ySAGA at 3.8–3.9 Å resolution provided insights into the histone-fold core, TBP binding and overall ySAGA architecture<sup>4,5</sup>, with the HAT and DUB domains being flexibly attached to the core. While vertebrate SAGA is highly conserved (roughly 95–58% sequence identity), the conservation with yeast drops dramatically (roughly 18% sequence identity) (Extended Data Fig. 1a and Supplementary Table 1), and numerous domain insertions, deletions and gene duplications have led to subfunctionalization of hSAGA subunits<sup>8</sup> and to hSAGA being essential for development in vertebrates (in contrast, ySAGA is not essential for viability)<sup>3,7</sup>. The compositional and functional differences between the yeast and human complexes hinted at possible structural differences and led us to examine the structure of hSAGA using cryogenic-electron microscopy (cryo-EM).

## Results

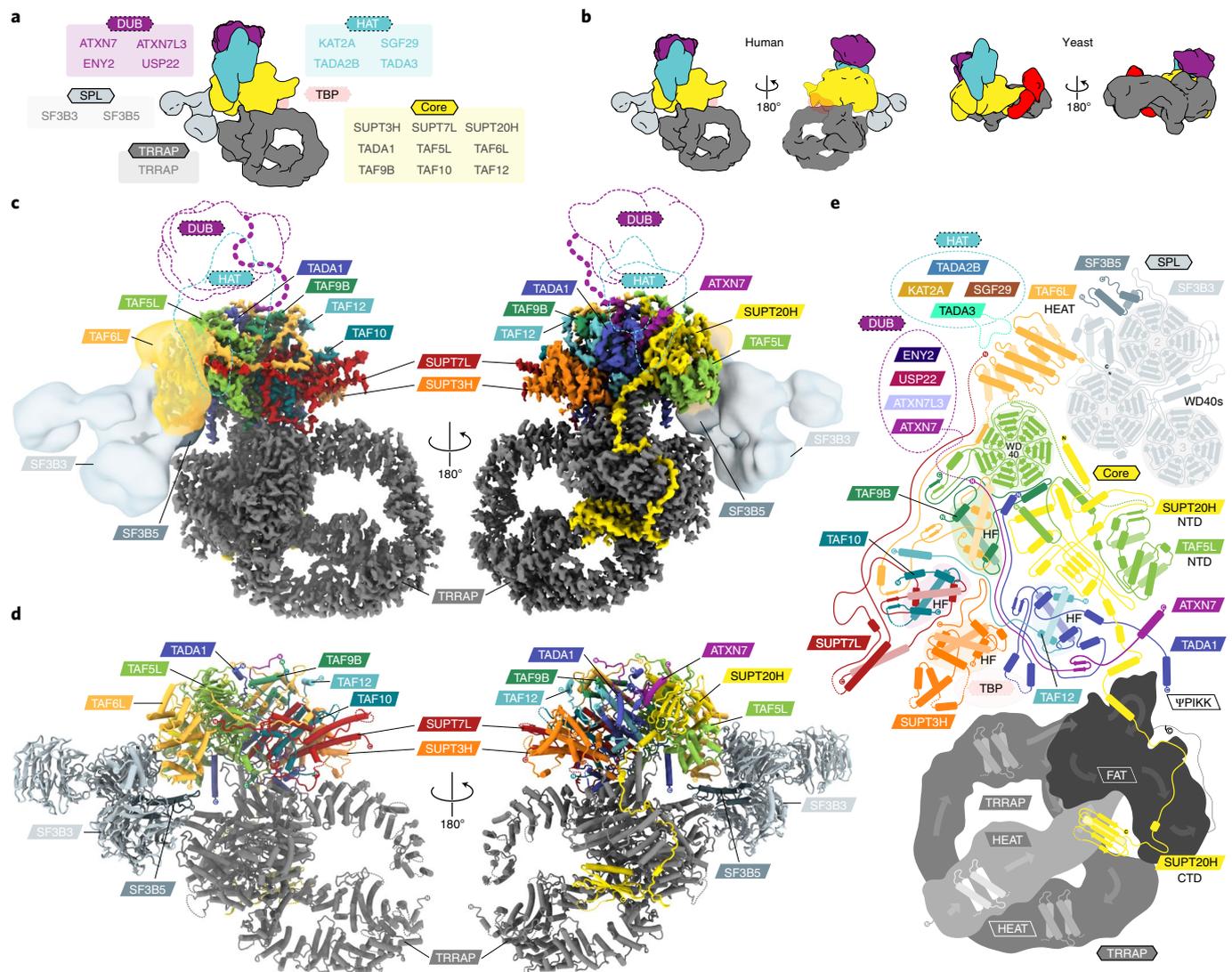
**Architecture of human SAGA.** For our structural studies, we purified intact, endogenous hSAGA from HeLa cells (Methods). The presence of all 20 hSAGA subunits was validated using western blotting and mass spectrometry (Extended Data Fig. 1b,c, Supplementary Table 2 and Methods). Single-particle negative stain (Table 1 and Extended Data Fig. 2) showed the presence of a central core domain sitting atop the distinct cradle-shaped TRRAP module, with a Y-shaped density flexibly tethered to the core domain proximal to the TRRAP cradle. Our negative stain reconstruction at 19 Å resolution (Extended Data Fig. 2d) already revealed clear architectural differences with respect to the ySAGA complexes<sup>4,5</sup> (Fig. 1a,b). Using cryo-EM we then obtained a reconstruction (overall resolution of 2.9 Å; Fig. 1c, Table 1 and Extended Data Fig. 3a–g) that allowed us to build an atomic model for the best ordered regions of hSAGA: the core module, consisting of TAF5L, SUPT20H, seven histone-fold-containing subunits (TAF6L, TAF9B, TAF10, TAF12, SUPT7L, TADA1 and SUPT3H) and the DUB anchor subunit ATXN7, and the large TRRAP subunit that consists of a circular HEAT (Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1) repeat cradle, FAT (FRAP, ATM and TRRAP) and pseudo-(Ψ)PIKK domains (Fig. 1d,e, Extended Data Fig. 3 and Supplementary Video 1).

Due to the flexible nature of the region connecting the Y-shaped density, this region could not be resolved in the high-resolution cryo-EM reconstruction (Extended Data Fig. 3i,j). However, by superposing the common elements with the negative stain structure and following the main chain density for TAF6L, we were able to unambiguously assign this region to the metazoan-specific SPL module (Fig. 1c,d). We were able to dock with high precision a homology model of the TAF6L HEAT domain<sup>9</sup> as well as the SF3B3/SF3B5 subunits of the SF3b crystal structure<sup>10</sup> into our 19 Å map (Extended Data Fig. 2e–h).

<sup>1</sup>California Institute for Quantitative Biology (QB3), University of California, Berkeley, CA, USA. <sup>2</sup>Molecular Biophysics and Integrated Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>3</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA.

<sup>4</sup>Biophysics Graduate Group, University of California, Berkeley, CA, USA. <sup>5</sup>Howard Hughes Medical Institute, University of California, Berkeley, CA, USA.

<sup>6</sup>Present address: Department of Biology, Johns Hopkins University, Baltimore, MD, USA. <sup>7</sup>Present address: School of Public Health, Sun Yat-sen University, Shenzhen, China. <sup>8</sup>These authors contributed equally: Dominik A. Herbst, Meagan N. Esbin. <sup>✉</sup>e-mail: [enogales@lbl.gov](mailto:enogales@lbl.gov)



**Fig. 1 | The modular architecture of hSAGA.** **a**, Schematic organization of hSAGA subunits into modules (boxes). The putative TBP binding region (based on ySAGA) is indicated as translucent red shape, because evidence for TBP binding by hSAGA is still lacking. **b**, Comparison of hSAGA (left) and ySAGA<sup>5</sup> (right) after superposition on the core module (Supplementary Fig. 1). Yeast modules are colored according to their human homolog. The yeast TBP binding module is shown in red. Viewing orientations in all other figures are indicated by these schemes. **c**, Hybrid map of hSAGA. The high-resolution cryo-EM map (core and TRRAP) (contoured at 7.0 Å) and the negative stain map for the TAF6L HEAT domain and the SPL module are shown (contoured at 6.1 Å). The expected DUB and HAT locations are indicated in dashed lines. Subunits are colored as indicated by the labels. **d**, Atomic model of hSAGA. **e**, Topology map of hSAGA subunits grouped by modules (not drawn to scale). Color schemes are consistent throughout all figures.

While all hSAGA HAT and DUB subunits were confirmed in our sample, they were not resolved in our structural analysis, either due to flexible tethering or a more dynamic and labile attachment, consistent with the low resolution and flexibility described for these modules in the yeast complex<sup>4,5</sup>. By comparing the positions of the HAT-tethering subunits TAF6L/SUPT7L and the integration of the DUB subunit ATXN7 in the core (Fig. 2a) with their respective counterparts in ySAGA, we anticipate similar general positions for these modules in hSAGA (Extended Data Fig. 4a–f). Moreover, very weak density, visible only in some class averages after gradient crosslinking (Extended Data Fig. 4g and Methods), likely correspond to the HAT and DUB domains, indicating a flexible or dynamic connection to the complex at the expected positions.

**Core module structure and tethering of the SPL module.** The structure of hSAGA is organized around the nine-subunit core module (Figs. 1a,c–e and 2a), in which the seven subunits contain

histone folds (SUPT3H containing two) assemble into a distorted pseudo-octamer (Fig. 2b,c), as also observed in ySAGA<sup>4,5</sup> (Extended Data Fig. 4h), as well as in human and yeast TFIID<sup>9,11,12</sup>. The distortion from the symmetric nucleosomal octamer creates a gap that is filled by the TAF5L WD40 propeller, which centrally binds to helix  $\alpha 2$  of the TAF6L histone fold (Fig. 2a,c). The periphery of the core is organized by the C-terminal TAF6L HEAT repeat domain, which connects the SPL module on its concave side (Figs. 1c and 2d) and probably the HAT module on its convex side (Extended Data Fig. 4a–c). Such connections are consistent with yeast two-hybrid assays of *Drosophila* homologs, which suggested interactions between SF3B3 and SF3B5 (SPL), SGF29 (HAT) and SUPT7L (Core)<sup>13</sup>. The SF3B3 subunit contains three WD40 propellers and tethers the SPL module via propeller one and two to the TAF6L HEAT repeats (Fig. 2d and Extended Data Fig. 2e,i). Of note, in ySAGA, the corresponding interface on the Taf6 HEAT repeat is blocked by the Taf5 N-terminal domain (NTD) (Fig. 2e). This

**Table 1 | Cryo-EM data collection and refinement statistics**

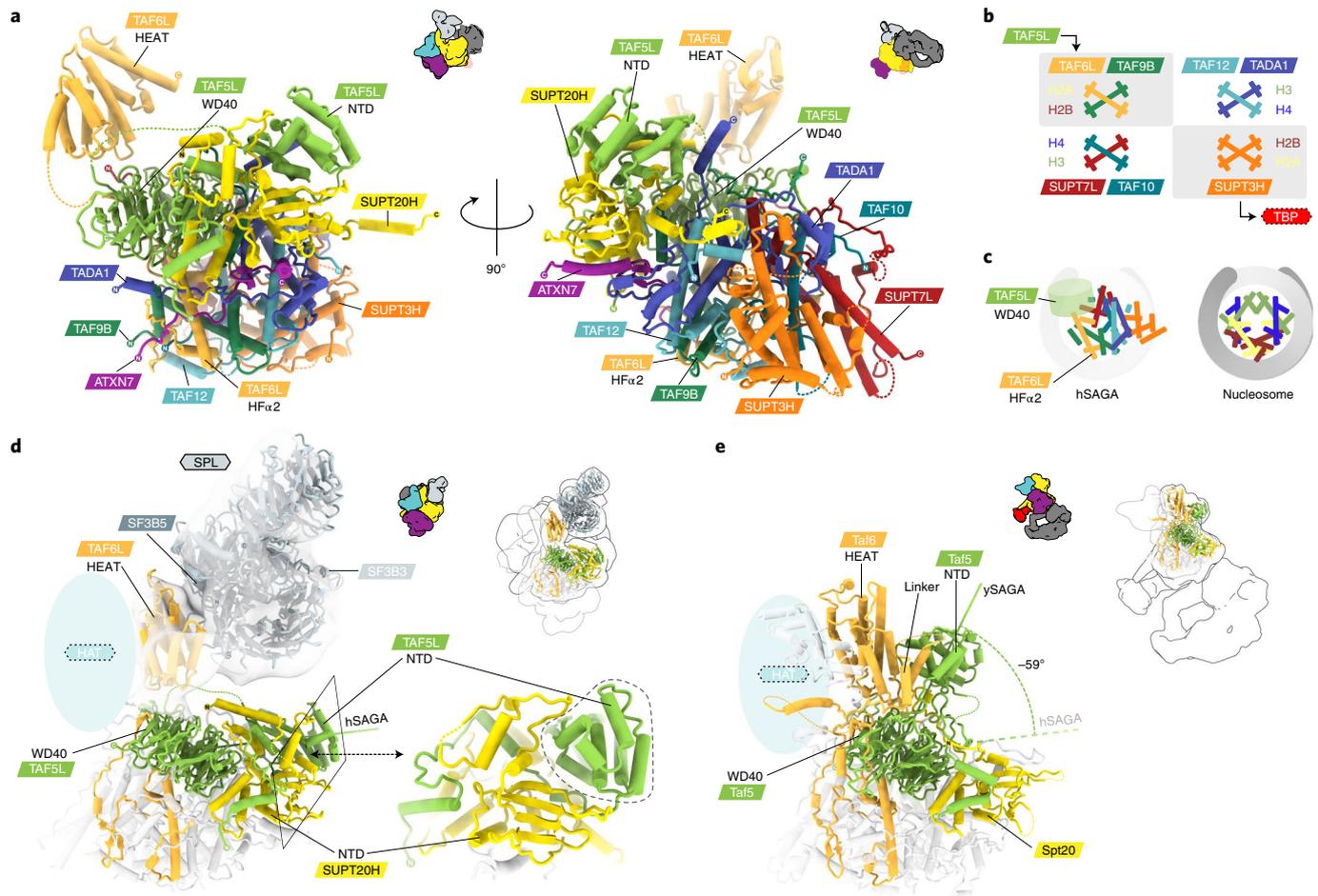
	Cryo-EM (EMDB-23027) (PDB 7KTR)	Negative stain (EMDB-23028) (PDB 7KTS)
<b>Data collection and processing</b>		
Microscope	FEI Titan Krios G2	FEI Technai F20
Camera	Gatan K3 Summit (super resolution)	Gatan UltraScan4000
Voltage (keV)	300	120
Magnification	64,000	80,000
Defocus range ( $\mu\text{m}$ )	0.9–3.4	0.4–3.9
Micrographs/videos	10,224	1
Frames per video	50	NA
Pixel size ( $\text{\AA}$ )	1.187	1.4
Total dose ( $\text{e}^-/\text{\AA}^{-2}$ )	50	35
Symmetry imposed	C1	C1
Particles initial/final	3,167,367/357,441	47,790/3,157
Map resolution ( $\text{\AA}$ )	2.9	19
FSC threshold	0.143	0.143
Map resolution range ( $\text{\AA}$ )	2.5–9.0	NA
<b>Model refinements</b>		
Initial model used (PDB code)	6F3T, de novo	7KTR, 5IFE
Method	real space, adp	rigid body <sup>a</sup>
Model resolution ( $\text{\AA}$ )	3.0	NA
FSC threshold	0.5	NA
Model resolution range ( $\text{\AA}$ )	2.3–25.4	NA
Map-to-model cross-correlation	0.80	0.52
Map sharpening $B$ factor ( $\text{\AA}^2$ )	NA <sup>b</sup>	–1,200
Model composition		
Nonhydrogen atoms	40,337	51,173
Protein residues	5,169	6,632
Ligands	1	0
Mean model $B$ factors ( $\text{\AA}^2$ )		
Protein	73.6	not refined
Ligand	66.6	not refined
R.m.s. deviations		
Bond lengths ( $\text{\AA}$ )	0.004	0.005
Bond angles ( $^\circ$ )	0.62	0.99
Validation		
MolProbity score	1.50	1.62
All-atom clashscore	3.42	4.79
Rotamer outliers (%)	0.02	0.02
C-beta deviations	0	0
Ramachandran plot		
Favored (%)	94.72	94.57
Allowed (%)	5.24	5.38
Disallowed (%)	0.04	0.05
Rama-Z score, whole (r.m.s.)	0.16 (0.12)	–0.35 (0.11)

Cryo-EM data collection and refinement statistics. <sup>a</sup>Three bodies were fit comprising the SF3B3/SF3B5 subunits of SF3b (PDB 5IFE), a homology model based on human TAF6 (PDB 6MZL) and the cryo-EM structure of hSAGA. <sup>b</sup>LocSpiral filtered: NA, not applicable; adp, atomic displacement parameters.

domain is rotated  $-59^\circ$  relative to the human TAF5L NTD, which in hSAGA is latched in place by the SUPT20H NTD (Fig. 2d).

**SUPT20H as a latch and binding of InsP<sub>6</sub>.** SUPT20H forms the largest interface with the rest of the complex (approximately

12,000  $\text{\AA}^2$ ) and acts as a clamp-like scaffold within hSAGA (Fig. 3a), supporting its central role in complex assembly and module association<sup>14,15</sup>. Our structure shows how SUPT20H tethers the DUB anchor ATXN7 to the core (Fig. 2a and Extended Data Fig. 4d). In addition to its crucial role in latching away the TAF5L NTD,



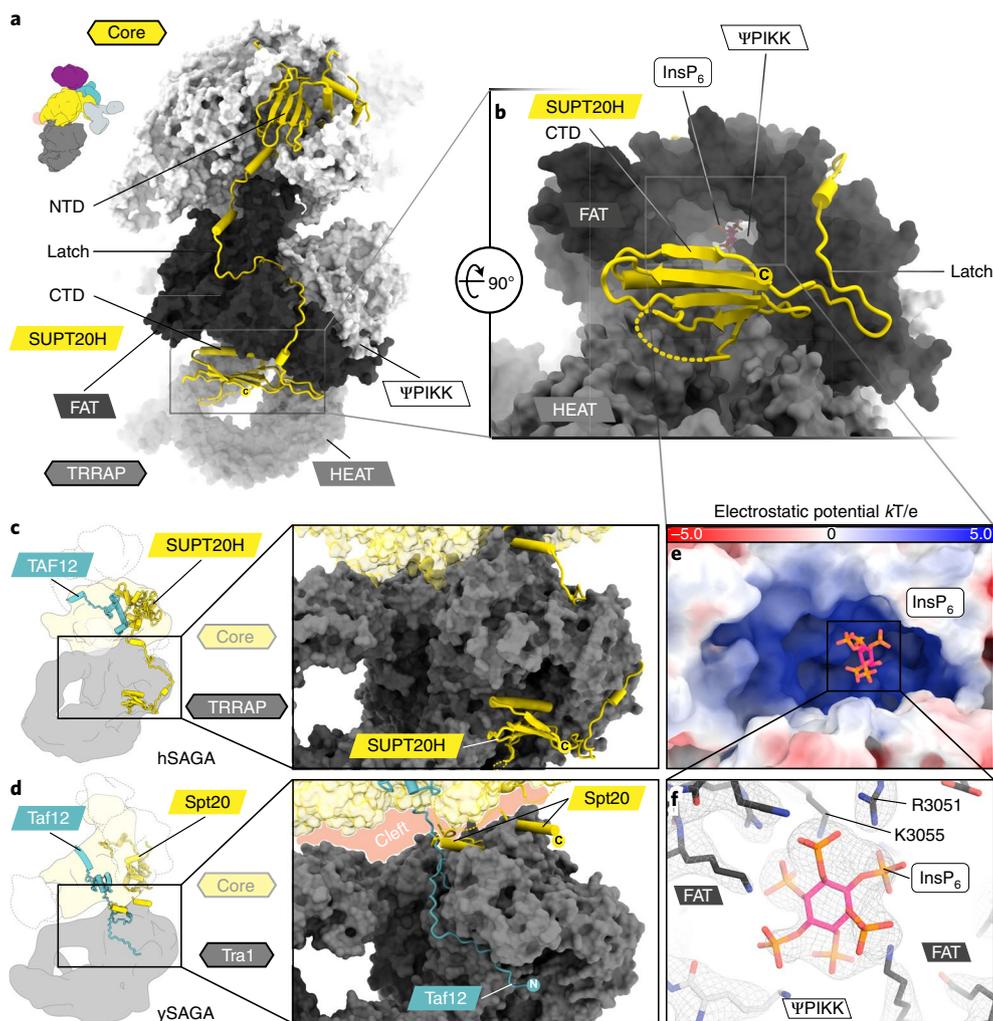
**Fig. 2 | Organization of the core and integration of the SPL module. a**, Atomic model of hSAGA core. **b**, Schematic comparison of the histone-fold core organization in hSAGA with the nucleosome. Histone-fold dimers are grouped in boxes, colored by hSAGA subunits (above/below). Corresponding histones are indicated on the side, as well as the contact with the TAF5L WD40 propeller and the potential interaction with TBP below/above. **c**, Schematic of the relative locations of the distorted histone-fold octamer helices in the hSAGA core and in the nucleosome. **d**, The (concave) surface of the TAF6L HEAT domain tethers the SPL module in hSAGA (translucent EM contoured at  $6.1\sigma$ ). The SUPT20H NTD latches the TAF5L NTD in place (a close-up on the right). **e**, In ySAGA, the Taf5 NTD is rotated  $-59^\circ$  and occupies the corresponding Taf6 HEAT domain surface (PDB 6T9K). The location of the HAT modules in **d** and **e** is indicated by translucent blue ovals and location of the depicted regions in the context of the complete complexes are outlined at the top right.

thus allowing incorporation of the SPL module (described above), SUPT20H also makes extensive contacts between the core and TRRAP module that contribute to create an architecture very different from that of ySAGA. The SUPT20H NTD connects to a long linker, ‘the latch’ (Fig. 3a), missing in yeast, that wraps along the surface of the core, around the TRRAP FAT domain and terminates in the cleft below the FAT and central TRRAP HEAT repeats with a previously unpredicted C-terminal domain (CTD) (Fig. 3b). The CTD folds into a five-stranded antiparallel beta-sheet with an alpha-helix parallel to the sheet that connects the two C-terminal outer strands (Figs. 1c–e and 3b). The closest structural homolog is the Spt6 SH2 domain of *Candida glabrata*<sup>16</sup> (2.20 Å  $C_\alpha$ -r.m.s.d. over 49 residues). Neither the SUPT20H latch, nor its CTD are conserved in ySAGA (Fig. 3c,d and Extended Data Fig. 5a,b). On the other hand, the N terminus of ySAGA Taf12, lacking in the human homolog, emerges from a location similar to that of the SUPT20H CTD and wraps around the opposite side of the Tra1 FAT domain (Fig. 3c,d). Metazoan TAF12s have a much shorter N terminus and contact TRRAP at a different location (Extended Data Fig. 5c,d).

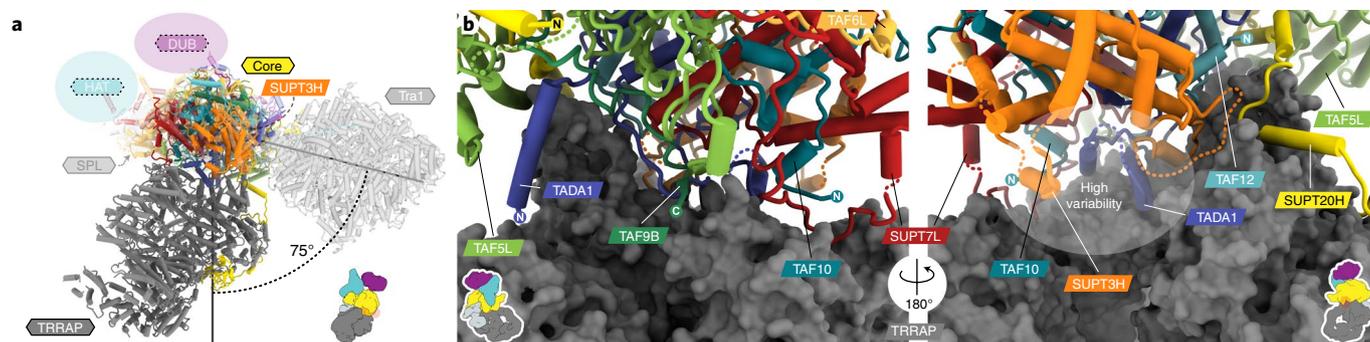
The CTD location of SUPT20H resembles a lid at the entrance of a positively charged tunnel below the FAT domain that is

conserved in metazoans (Fig. 3e and Extended Data Fig. 6a–d). In a side pocket of this tunnel and bound to highly conserved residues of the FAT and  $\Psi$ PIKK domains, our structure shows clear density for the metabolite inositol hexakisphosphate ( $\text{InsP}_6$ ), which copurified with hSAGA (Fig. 3b,e,f and Extended Data Fig. 6e–j).

**TRRAP structure and interactions with the core module.** The TRRAP subunit, like the yeast Tra1, has a characteristic tripartite HEAT repeat organization, consisting of a central N-terminal repeat and a circular cradle, followed by a FAT domain and a  $\Psi$ PIKK (Fig. 1c–e) (the Tra1 and TRRAP subunits are shared with the yeast NuA4 complex and its human counterpart, TIP60, respectively)<sup>4–6,17</sup>. Compared to ySAGA, hSAGA exhibits a dramatically different TRRAP–core interface that leads to a relative rotation of  $75^\circ$  of TRRAP/Tra1 with respect to the core and SUPT3H/Spt3 (Fig. 4a). While the approximate region of the interface is similar on TRRAP and Tra1, the region on the core contributing to the interface is dramatically different for yeast and human complexes. In hSAGA, all core subunits except for TAF6L and ATXN7 are involved in the TRRAP–core interface (Fig. 4b and Extended Data Fig. 7a–c), as compared to a limited number in yeast (Extended Data



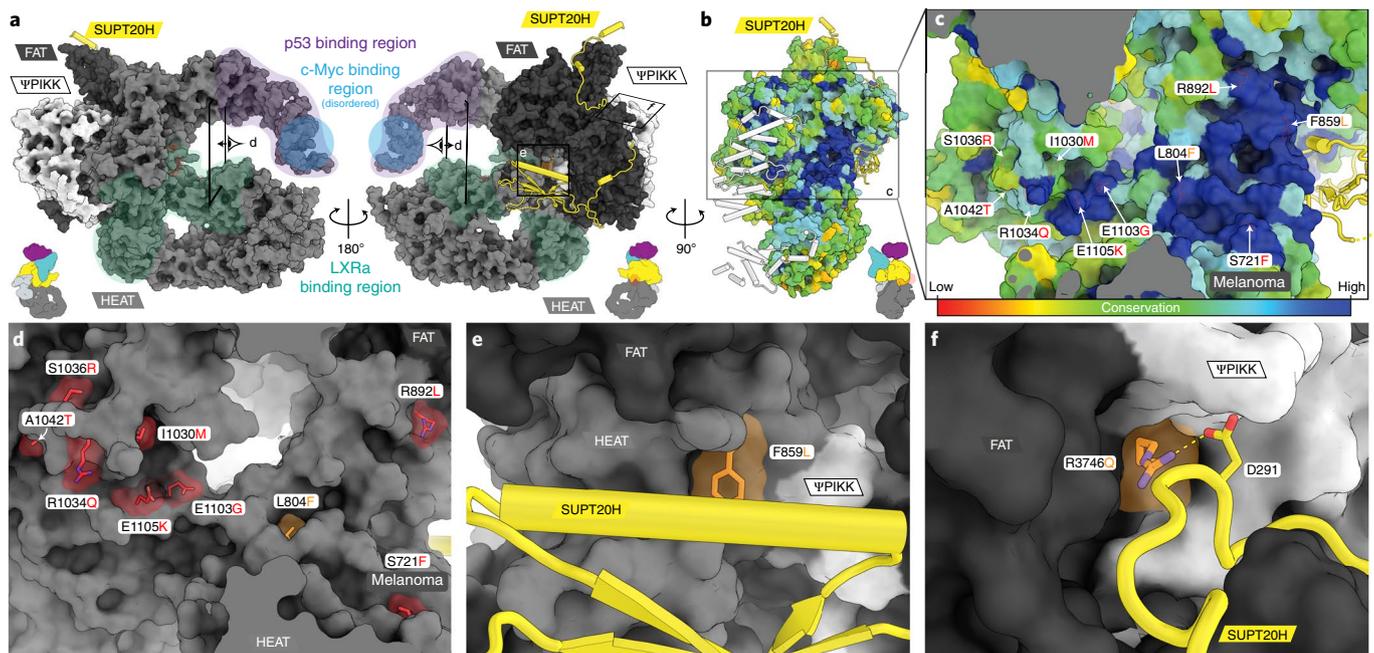
**Fig. 3 | SUPT20H structure and interactions.** **a**, SUPT20H contains an NTD in the core and a CTD in the TRRAP module, that are connected via a linker ('latch') that runs along a surface groove around the TRRAP FAT domain. TRRAP domains are colored in different shades of gray. **b**, The CTD is located at the entrance to a positively charged tunnel below the FAT domain. **c,d**, Schematic and close-up view of hSAGA (**c**) and ySAGA (**d**, PDB 6T9I), showing the different interactions of SUPT20H/Spt20 and TAF12/Taf12 with TRRAP/Tra1. The cleft between the Tra1 and core module is indicated in light red (**d**). **e**, Electrostatic surface potential representation of the positively charged tunnel (**b**) with InsP<sub>6</sub> (**f**) (Extended Data Fig. 6). **f**, LocSpiral filtered multibody map (contoured at 11 $\sigma$ ) showing the InsP<sub>6</sub> site. Carbon atoms are colored by TRRAP domain or pink for InsP<sub>6</sub>, red (oxygen), blue (nitrogen) or orange (phosphorus).



**Fig. 4 | Distinct TRRAP tethering in hSAGA.** **a**, Superposition of the hSAGA and ySAGA (PDB 6T9I) (translucent) models on the core. Other module locations are indicated by colored ovals. **b**, Close-up views of the hSAGA TRRAP-core interface. A region of high variability is indicated (Extended Data Fig. 8b,c).

Fig. 7d–i). In ySAGA, the core subunits Spt20 and Taf12 form local interactions on the Tra1 surface and are connected to the core by flexible linkers that span a large cleft between the modules (Fig. 3d

and Extended Data Fig. 7f,i). Such a cleft does not exist in hSAGA (Extended Data Fig. 7c) and presumably leads to the increased flexibility observed between the core and Tra1 in yeast<sup>18,19</sup>. While the



**Fig. 5 | Activator binding and mapping of disease mutations in TRRAP.** **a**, Surface representation of TRRAP colored by domains. Mapped activator-binding regions are indicated as colored overlays. The c-Myc binding region is probably located in a disordered loop between two helices. Boxes indicate the relative views in the following panels. **b**, Surface representation colored by conservation (for details, see Extended Data Fig. 6c). **c,d** Residue mutations associated with cancer, autism, or intellectual disability<sup>34,37</sup>. **c**, Most disease mutations lie in a region of high sequence conservation. Surface coloring as in **b**. **d**, Surface representation of disease mutations as shown in **a**. Red, surface exposed, probably interfere with activator binding; orange, buried, likely to structurally destabilize TRRAP (Extended Data Fig. 10a,b) or the interaction with SUPT20H (**e,f**). **e**, A disease-causing mutation of F859 is located at the interface with the SUPT20H CTD. **f**, R3746 forms a salt bridge with D291 of the SUPT20H latch. The disease mutation R3746Q disrupts the salt bridge. The reported residue numbers relate to the modeled isoform (Uniprot F2Z2U4).

main TRRAP–/Tra1–core interfaces, corresponding to the core’s footprint on TRRAP or Tra1 (Fig. 4b and Extended Data Fig. 7b,c,e,f,h,i), are of a similar size (roughly 3,500 Å<sup>2</sup>), both complexes rely on additional stabilization by unique extensions of either Taf12 in yeast (Fig. 3d, Extended Data Figs. 5c and 7f,i and Supplementary Table 3) or SUPT20H in human that form interfaces with different regions on Tra1 and TRRAP, respectively (Fig. 3c and Extended Data Figs. 5a and 7c). In hSAGA, the SUPT20H extension doubles the total interface (to 7,073 Å<sup>2</sup>), which is ultimately 64% larger than that of ySAGA (Supplementary Table 3).

## Discussion

**Local variations in the core enable a divergent architecture.** Our study revealed that local variations, such as the repositioning of the TAF5L NTD and different interactions of SUPT20H and TAF12 on the TRRAP surface, result in very different interfaces between the structurally conserved cores of ySAGA and hSAGA with the Tra1 and TRRAP subunits, respectively. Consequently, this non-conserved geometry positions functional elements in the core and the activator-binding subunit in totally different relative orientations. While the hSAGA TRRAP–core interface is not entirely rigid (Extended Data Fig. 8a), a potential transition between the observed yeast and human conformations, is extremely unlikely. The yeast Taf12 N terminus and Spt20 C terminus form local interactions on the surface of Tra1 beyond the cleft and are likely to move with it as one body. Rearrangement from the yeast to the human conformation would require unfolding of these elements on Tra1 or of parts of the Taf12 histone fold. Similarly, a transition from the human to the yeast conformation would require unfolding of SUPT20H NTD elements that are involved in TAF5L NTD binding. Such a transition far exceeds the conformational space that these modules appear to be capable of exploring. Within the NuA4 complex, Tra1

has been shown to connect to the rest of the complex using a similar, albeit larger interface region as in ySAGA<sup>20</sup>, suggesting that the newly defined TRRAP interface in hSAGA might also be relevant for TRRAP incorporation into the related metazoan TIP60 complex.

**Human SAGA and TBP.** Analysis of our cryo-EM data (Methods) revealed heterogeneity that suggests alternative main chain conformations in the cleft between TRRAP and SUPT3H/SUPT7L/TADA1, which includes the region where TBP is bound by Spt3 (SUPT3H homolog) and the yeast-specific Spt8 (ref. 5) (Fig. 1e and Extended Data Figs. 3i,j and 8). We could not observe density for TBP, even when it was added in excess to the purified hSAGA (Methods), in contrast with the observations for the yeast complex<sup>3</sup>, highlighting another distinct difference between these complexes. The lack of a stable TBP–hSAGA complex may either indicate that hSAGA does not bind TBP at all, or, together with the observed electron microscopy (EM) heterogeneity, might indicate a highly dynamic or regulated mode of TBP binding, unlike that for the TFIID or ySAGA complex, that may require stabilization by additional factors. Metazoans lack a homolog of the yeast subunit Spt8, which is sufficient for TBP binding on its own, whereas Spt3 is not<sup>21</sup>. On the other hand, the transcription factor c-Myc has been shown to interact with TBP<sup>22</sup> as well as TRRAP<sup>23,24</sup> via nonoverlapping regions, suggesting the intriguing possibility that activators could play a role in TBP recruitment to metazoan SAGA. DNA binding by ySAGA-bound TBP was shown to be sterically hindered by Tra1 (ref. 5). However, due to the distinct tethering of TRRAP in hSAGA, any interaction of TBP with hSAGA could have different consequences on TBP–DNA binding.

**Metazoan incorporation of a SPL module.** Comparison of our structure with that of ySAGA reveals a crucial rearrangement of the TAF5L NTD within the core. The lack of a stabilizing TAF6L

HEAT-TAF5L NTD interaction probably contributes to increased flexibility of the TAF6L HEAT repeat domain, a critical platform for HAT and SPL module integration. Furthermore, the local repositioning of the TAF5L NTD exposes the TAF6L interface to allow for SPL module incorporation in hSAGA. The position of the TAF5 NTD is also dramatically different from that observed in Lobe A and Lobe B of TFIID, making this domain a crucial marker for the divergent architectures of TAF-containing complexes<sup>9,11</sup> (Extended Data Fig. 4i). While our EM structure revealed the site of incorporation of the SPL module, very little is yet known about its function or how its components partition between SAGA and the U2 small nuclear ribonucleoprotein. It has been proposed that SF3B3 incorporation into hSAGA may play a role in ultraviolet (UV)-damaged DNA binding and repair<sup>25</sup>, but contradictory results argued that the structurally related subunit DDB1, which we did not observe in our sample, is the one that recognizes UV-damaged DNA in the context of hSAGA (Supplementary Table 2)<sup>26</sup>. The SPL module subunits, SF3B3 and SF3B5, are shared with the metazoan spliceosomal SF3b core complex within the U2 small nuclear ribonucleoprotein. Our structure shows that they are tethered to the rest of the hSAGA complex in a similar way as they are in the spliceosomal SF3b complex<sup>27</sup>. In hSAGA, SF3B3 binds to the HEAT repeat domain of TAF6L, while SF3B3 binds to the HEAT solenoid of SF3B1 in the SF3b complex<sup>10</sup>, and they do so using an overlapping interface (Extended Data Fig. 2i,j). Therefore, SF3B3/SF3B5 cannot be simultaneously incorporated into hSAGA and the SF3b SPL complex.

**Pseudo-kinase active site in TRRAP.** TRRAP lacks kinase activity, although homologs of TRRAP are present in active kinases, such as mTOR, DNA-PKcs and ATM<sup>28</sup> (Extended Data Fig. 9a–e). While the SAGA ΨPIKK lacks the canonical active site residues for catalysis<sup>23,28</sup> (Extended Data Fig. 9f), we found that the first residue of the TRRAP activation loop (Y3698), corresponding to the aspartate in the active PIKK's DFG motif<sup>23</sup>, adopts an unusual and well defined *cis*-peptide bond (Extended Data Fig. 9g). Such geometry outliers often serve a function in active sites<sup>29</sup>, and its position in our structure, together with the high evolutionary conservation of the ΨPIKK and of this specific residue in metazoans (Extended Data Fig. 9f), raises the question of whether the inactive kinase might have a different and so far undiscovered function, as observed for other pseudokinases<sup>30</sup>.

**Binding of InsP<sub>6</sub> and its possible role in TRRAP stability.** The resolution of our structure allowed us to visualize InsP<sub>6</sub> in the positively charged pocket below the TRRAP FAT domain. The position of InsP<sub>6</sub> in hSAGA is equivalent to that observed for mTORC2 (ref. <sup>31</sup>) (Fig. 3b,e,f and Extended Data Fig. 6g,h) or the SMG1 kinase<sup>32</sup>, and thus it could serve a similar stabilizing role as proposed for those kinases<sup>31,32</sup>. In the ySAGA structures<sup>4,5</sup>, the region surrounding this pocket, including residues corresponding to R3051 and K3055 in hSAGA (Fig. 3f and Extended Data Fig. 6e), is poorly resolved and lacks InsP<sub>6</sub> density (Extended Data Fig. 6i). On the other hand, an earlier structure of the yeast Tra1 subunit<sup>17</sup> is better defined in this region and contains an unattributed density where InsP<sub>6</sub> is seen bound in hSAGA (Extended Data Fig. 6j), potentially linking the stability of the TRRAP FAT domain to the presence of InsP<sub>6</sub>.

**Human disease mutations.** The best characterized function of SAGA's TRRAP module is serving as an interaction hub for transcriptional activators, which leads to its critical role in many diseases and its consideration as a prognostic marker and therapeutic target in many cancers<sup>23,24,28,33–39</sup>. Structurally, TRRAP displays high flexibility around the N-terminal cradle region where the c-Myc and p53 binding sites are located<sup>24,36</sup> (Fig. 5a), and thus it is possible that c-Myc/p53 binding could stabilize or mediate

conformational changes in this region. A cluster of disease-causing mutations lies along a highly conserved FAT-proximal HEAT repeat region where the N-terminal HEAT repeat arm and circular cradle meet (Fig. 5b,c), a site that has been shown to be crucial for liver X receptor alpha (LXRα) interaction<sup>28,33,34,37</sup>. A number of mutations, including the prevalent melanoma mutation S722F (TRRAP isoform here, S721F), are part of a highly conserved surface patch and probably involved in effector binding (Fig. 5c,d). Other mutations appear buried and are likely to affect folding of the HEAT repeats and interfere with the structural integrity of TRRAP (Extended Data Fig. 10a,b). Two independent mutations identified in patients with intellectual disability and neurodevelopmental disorders<sup>37</sup> are at sites of interaction with the metazoan-specific extension seen in SUPT20H. The first (F859L) localizes directly at the interface with the SUPT20H CTD (Fig. 5e) and the second (R3746Q) eliminates a salt bridge with the highly conserved D291 of the SUPT20H latch (Fig. 5f and Extended Data Fig. 5b). Because TRRAP is a scaffold for other important cellular complexes, disease-causing mutations may also disrupt assembly or lead to perturbations within TIP60 (ref. <sup>28</sup>).

### Concluding remarks

Our hSAGA structure reveals conserved structural elements as well as notable divergences from the yeast complex, including a distinct architecture and TRRAP–core interface, a lack of stable interaction with TBP and the visualization of the incorporation of the metazoan SPL module. Human SAGA complex combines transcription factor-interacting and enzymatic modules that need to regulate an intricate and unique transcriptional and chromatin landscape within human cells, in which enhancers and promoters are separated by kilo- to megabase distances. Furthermore, human promoter architectures, as well as intron and splice site properties, are very distinct from those in yeast<sup>40,41</sup>. These newly revealed structural features of hSAGA probably reflect unique mechanisms for this complex in human cells that go beyond transcription and chromatin regulation and can provide a launching point for further studies of SAGA's roles in human disease.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-021-00682-7>.

Received: 27 July 2021; Accepted: 5 October 2021;

Published online: 22 November 2021

### References

1. Timmers, H. T. M. SAGA and TFIID: friends of TBP drifting apart. *Biochim. Biophys. Acta Gene Regul. Mech.* **1864**, <https://doi.org/10.1016/j.bbagr.2020.194604> (2021).
2. Fischer, V., Schumacher, K., Tora, L. & Devys, D. Global role for coactivator complexes in RNA polymerase II transcription. *Transcription* **10**, 29–36 (2019).
3. Wang, L. & Dent, S. Y. Functions of SAGA in development and disease. *Epigenomics* **6**, 329–339 (2014).
4. Wang, H. et al. Structure of the transcription coactivator SAGA. *Nature* **577**, 717–720 (2020).
5. Papai, G. et al. Structure of SAGA and mechanism of TBP deposition on gene promoters. *Nature* **577**, 711–716 (2020).
6. Helmlinger, D. & Tora, L. Sharing the SAGA. *Trends Biochem. Sci.* **42**, 850–861 (2017).
7. Cheon, Y., Kim, H., Park, K., Kim, M. & Lee, D. Dynamic modules of the coactivator SAGA in eukaryotic transcription. *Exp. Mol. Med.* **52**, 991–1003 (2020).
8. Antonova, S. V., Boeren, J., Timmers, H. T. M. & Snel, B. Epigenetics and transcription regulation during eukaryotic diversification: the saga of TFIID. *Gene Dev.* **33**, <https://doi.org/10.1101/gad.300475.117> (2019).

9. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, <https://doi.org/10.1126/science.aau8872> (2018).
10. Cretu, C. et al. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016).
11. Kolesnikova, O. et al. Molecular structure of promoter-bound yeast TFIID. *Nat. Commun.* **9**, 4666 (2018).
12. Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, <https://doi.org/10.1126/science.aba8490> (2021).
13. Stegeman, R. et al. The spliceosomal protein SF3B5 is a novel component of *Drosophila* SAGA that functions in gene expression independent of splicing. *J. Mol. Biol.* **428**, 3632–3649 (2016).
14. Elias-Villalobos, A., Toulecc, D., Faux, C., Seveno, M. & Helmlinger, D. Chaperone-mediated ordered assembly of the SAGA and NuA4 transcription co-activator complexes in yeast. *Nat. Commun.* **10**, 5237 (2019).
15. Nagy, Z. et al. The Human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. *Mol. Cell. Biol.* **29**, 1649–1660 (2009).
16. Dengl, S., Mayer, A., Sun, M. & Cramer, P. Structure and in vivo requirement of the yeast Spt6 SH2 domain. *J. Mol. Biol.* **389**, 211–225 (2009).
17. Diaz-Santín, L. M., Lukyanova, N., Aciyan, E. & Cheung, A. C. Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. *eLife* **6**, <https://doi.org/10.7554/eLife.28384> (2017).
18. Sharov, G. et al. Structure of the transcription activator target Tra1 within the chromatin modifying complex SAGA. *Nat. Commun.* **8**, 1556 (2017).
19. Setiawati, D. et al. Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. *J. Biol. Chem.* **290**, 10057–10070 (2015).
20. Wang, X., Ahmad, S., Zhang, Z., Cote, J. & Cai, G. Architecture of the *Saccharomyces cerevisiae* NuA4/TIP60 complex. *Nat. Commun.* **9**, 1147 (2018).
21. Sermwittayawong, D. & Tan, S. SAGA binds TBP via its Spt8 subunit in competition with DNA: implications for TBP recruitment. *EMBO J.* **25**, 3791–3800 (2006).
22. Wei, Y. et al. Multiple direct interactions of TBP with the MYC oncoprotein. *Nat. Struct. Mol. Biol.* **26**, 1035–1043 (2019).
23. McMahon, S. B., Van Buskirk, H. A., Dugan, K. A., Copeland, T. D. & Cole, M. D. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **94**, 363–374 (1998).
24. Feris, E. J., Hinds, J. W. & Cole, M. D. Formation of a structurally-stable conformation by the intrinsically disordered MYC:TRRAP complex. *PLoS ONE* **14**, e0225784 (2019).
25. Brand, M. et al. UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. *EMBO J.* **20**, 3187–3196 (2001).
26. Martinez, E. et al. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol. Cell. Biol.* <https://doi.org/10.1128/mcb.21.20.6782-6795.2001> (2002).
27. Sun, C. The SF3b complex: splicing and beyond. *Cell Mol. Life Sci.* **77**, 3583–3595 (2020).
28. Murr, R., Vaissiere, T., Sawan, C., Shukla, V. & Herceg, Z. Orchestration of chromatin-based processes: mind the TRRAP. *Oncogene* **26**, 5358–5372 (2007).
29. Weiss, M. S., Jabs, A. & Hilgenfeld, R. Peptide bonds revisited. *Nat. Struct. Mol. Biol.* **5**, 676–676 (1998).
30. Reiterer, V., Evers, P. A. & Farhan, H. Day of the dead: pseudokinases and pseudophosphatases in physiology and disease. *Trends Cell Biol.* **24**, 489–505 (2014).
31. Scaiola, A. et al. The 3.2-Å resolution structure of human mTORC2. *Sci. Adv.* **6**, <https://doi.org/10.1126/sciadv.abc1251> (2020).
32. Gat, Y. et al. InsP6 binding to PIKK kinases revealed by the cryo-EM structure of an SMG1-SMG8-SMG9 complex. *Nat. Struct. Mol. Biol.* **26**, 1089–1093 (2019).
33. Unno, A. et al. TRRAP as a hepatic coactivator of LXR and FXR function. *Biochem. Biophys. Res. Commun.* **327**, 933–938 (2005).
34. Wei, X. et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat. Genet.* **43**, 442–446 (2011).
35. Wang, J. et al. Analysis of TRRAP as a potential molecular marker and therapeutic target for breast cancer. *J. Breast Cancer* **19**, 61–67 (2016).
36. Ard, P. G. et al. Transcriptional regulation of the mdm2 oncogene by p53 requires TRRAP acetyltransferase complexes. *Mol. Cell Biol.* **22**, 5650–5661 (2002).
37. Cogne, B. et al. Missense variants in the histone acetyltransferase complex component gene TRRAP cause autism and syndromic intellectual disability. *Am. J. Hum. Genet.* **104**, 530–541 (2019).
38. McMahon, S. B., Wood, M. A. & Cole, M. D. The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Mol. Cell Biol.* **20**, 556–562 (2000).
39. Herceg, Z. et al. Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. *Nat. Genet.* **29**, 206–211 (2001).
40. Dobi, K. C. & Winston, F. Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **27**, 5575–5586 (2007).
41. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## Methods

**SUPT7L-Halo-(FLAG)<sub>3</sub> knock-in cell line generation.** Human HeLa cells were cultured at 37 °C and 5% CO<sub>2</sub> in 4.5 g l<sup>-1</sup> glucose DMEM supplemented with 10% fetal bovine serum and 10 U ml<sup>-1</sup> penicillin-streptomycin, and subcultured at a ratio of 1:3 to 1:8 every 2 to 4 d. Genome editing was performed as described previously<sup>42</sup>. Wild-type HeLa cells were cotransfected with a Cas9 plasmid (CBh-driven 3xFLAGSV40NLS-pSpCas9-NLS; PGK-driven mVenus; U6-driven single-guide (sg) RNA) and a repair plasmid containing Halo-(FLAG)<sub>3</sub> flanked by roughly 800 bp of genomic homology sequence to *SUPT7L* on either side (18 µg of repair vector and 6 µg of Cas9 vector per P100 dish; 1:3 w/w) using Lipofectamine 2000 (Thermo Fisher catalog no. 11668019) according to the manufacturer's protocol. Four sgRNAs were designed using the Zhang laboratory CRISPR design tool (<https://zlab.bio/guide-design-resources>), cloned into the Cas9 vector and cotransfected with the repair vector individually. After 18–24 h, transfected cells were combined and sorted using fluorescence activated cell sorting for mVenus fluorescence. mVenus-sorted cells were grown for 4–12 d, labeled with 500 nM Halo-TMR and cell populations with higher fluorescence than TMR-labeled wild-type cells were fluorescence activated cell sort-selected and sorted individually into 96-well plates. Clones were expanded and genotyped by PCR. Successfully edited clones were further verified by PCR using multiple primer combinations, Sanger sequencing and western blotting.

**Preparative HeLa cell culture and nuclei extraction.** Large scale cultures of SUPT7L-Halo-(FLAG)<sub>3</sub> HeLa cells were grown at 37 °C and ambient CO<sub>2</sub> in a Hotpack Environmental Chamber (Scientific Products) in Joklik-modified Minimum Essential Medium Eagle (Sigma) media supplemented with 5% bovine calf serum, 50 U of penicillin-streptomycin and 2 mM Glutamax (Thermo Fisher). Cells were maintained in 61 Florence round-bottom spinning flasks (Fisher Scientific) each containing 41 of HeLa cultures and constantly stirred via a Precision Magnetic Stirrer Platform (Bellco). Every 24 h, cells were split 1:2 into fresh media grown to a density of roughly 2.5–5 × 10<sup>5</sup> cells per ml and collected. To collect, SUPT7L-Halo-(FLAG)<sub>3</sub> HeLa cells were centrifuged using a Fiberlite F9-6 × 1,000 LEX Fixed-Angle Rotor (Thermo Fisher) at 4 °C and 4,000 r.p.m. for 15 min. Cells were washed in PBSM (PBS buffer with 5 mM MgCl<sub>2</sub>) then centrifuged using an Eppendorf A-4-62 Swinging Bucket Rotor at 3,800 r.p.m. for 10 min. Cells were resuspended in 5 volumes of buffer A (10 mM HEPES pH 7.6, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 × Roche cComplete protease inhibitors) briefly vortexed, incubated on ice for 20 min and centrifuged (Eppendorf A-4-62, 3,800 r.p.m., 10 min, 4 °C). Cells were lysed by resuspension in 2 volumes of buffer A and dounced seven times using a glass homogenizer with a type B pestle. Nuclei were pelleted by centrifugation (Eppendorf A-4-62, 2,700 r.p.m., 10 min, 4 °C), flash frozen in liquid nitrogen and stored at –80 °C until use.

**hSAGA purification.** All steps were performed at 4 °C. Frozen nuclei from roughly 30 to 401 of cell culture were thawed, 0.9 volumes of buffer C (20 mM HEPES pH 7.8, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 25% glycerol, 0.42 M KCl, 1 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and 1 µM Leupeptin) added and dounced using a glass homogenizer and a type B pestle 20 times on ice. The nuclear extract was then centrifuged using a JA-20 Beckman rotor at 4 °C and 20,000 r.p.m. for 30 min. The supernatant was collected and adjusted to a conductivity of 0.3 M NaCl. The nuclear extract (roughly 60 ml) was loaded onto a 50 ml phosphocellulose P11 (GE Healthcare/Whatman) column, washed with 3 column volumes (CVs) of 0.3 M NaCl HEMG (20 mM HEPES-KOH pH 7.6, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 10% glycerol, 1 mM DTT, 0.5 mM PMSF and 1 µM Leupeptin), then eluted in two steps with 3 CVs 0.5 M NaCl HEMG, followed by 3 CVs of 1.0 M NaCl HEMG and fractionated (5 ml). Peak fractions were determined by Bradford assay and combined. Human SAGA eluted with the 0.5 M NaCl HEMG peak (hereafter called P0.5M) and dialyzed against 150 mM KCl buffer D (20 mM HEPES pH 7.8, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 10% glycerol, 150 mM KCl, 0.5 mM PMSF and 1 µM leupeptin) using SnakeSkin 10 kDa molecular weight cutoff dialysis tubing (Thermo Fisher). The dialyzed P0.5M fraction was supplemented with IGEPAL CA-630 (0.1% (v/v) final) and incubated with 500 µl of beads of FLAG M2 resin (Sigma) for 12 h nutating. The resin was washed twice with 2 CV of Column Buffer (25 mM HEPES pH 7.8, 0.2 M NaCl, 10% (v/v) glycerol, 1 mM EDTA, 0.5 mM TCEP, 0.1% (v/v) IGEPAL CA-630, 1 × Roche cComplete protease inhibitors), twice with 2 CV of Wash Buffer (Column buffer containing 0.6 M NaCl) and twice with 2 CV of Column Buffer. To elute, the beads were incubated with Column Buffer with 0.1 mg ml<sup>-1</sup> 3xFLAG peptide rocked for 1 h, then centrifuged (Eppendorf 022653041 fixed-angle rotor, 3,300 r.p.m., 5 min) and this was repeated for four 1-h elutions. Elutions were concentrated fivefold using a 100 kDa molecular weight cutoff Spin-X UF concentrator (Corning). The sample was frozen in liquid nitrogen and stored at –80 °C. Sample quality and the effect of freeze-thaw cycles were analyzed by negative stain EM, and elution fractions yielded a similar quality in cryo-EM. A concentration of approximately 50 nM was determined by densitometry.

**Antibody information.** The following primary antibodies were purchased from commercial suppliers and used at the indicated dilutions for western blotting. Anti-SUPT7L catalog no. sc-514548 (1:1,000) and anti-USP22 no. sc-390585

(1:200) were purchased from Santa Cruz Biotechnology. Anti-KAT2A catalog no. 3305 (1:1,000) was purchased from Cell Signaling Technology. Anti-TADA2B catalog no. PA5-57393 (1:2,500) was purchased from Thermo Fisher Scientific. Anti-TBP no. ab51841 (1:2,000) and anti-ENY2 no. ab183622 (1:1,000) were purchased from Abcam. Anti-TAF10 no. MABE1079 (1:2,000) was purchased from Millipore Sigma. Anti-TAF9B no. G2306 (1:500) is a homemade antibody previously created and validated in ref.<sup>43</sup>.

**Mass spectrometry.** Samples of purified hSAGA (roughly 1 µg) were shipped and analyzed by mass spectrometry by the Whitehead Institute Proteomics Facility (Cambridge, MA). Samples were diluted to 100 µl in 6 M urea, 100 mM Tris pH 7.8 buffer. Dithiothreitol (DTT, 5 µl of 200 mM) was added and incubated for 60 min at room temperature. Cysteines were alkylated by addition of 20 µl of 200 mM iodoacetamide and incubated for 60 min at room temperature. The sample was diluted to 900 µl with 100 mM Tris pH 7.8. The protein was digested by adding 100 µl of a 20 ng µl<sup>-1</sup> trypsin or chymotrypsin solution and incubated overnight at 37 °C. The resulting peptides were washed, extracted and concentrated by solid phase extraction using Waters Sep-Pak Plus C18 cartridges. Organic solvent was removed and volumes reduced to 15 µl by SpeedVac at 60 °C. The extracts were analyzed by reversed phase high-performance liquid chromatography using Waters NanoAcquity pumps and autosampler along with a Thermo Fisher Orbitrap Elite mass spectrometer using a nano flow configuration operated in a data dependent manner for 60 min. Fragmentation spectra were correlated against the Uniprot isoforms and TrEMBL databases for *Homo sapiens* using Sequest (Thermo Fisher Scientific; IseNode in Proteome Discoverer v.1.4.1.14). Sequest was searched (ion mass tolerance, 0.50 Da; parent ion tolerance, 15 ppm) with carbamidomethyl cysteine as fixed and methionine oxidation as variable modification. Consensus reports were obtained using Scaffold v.4.11.0 (Proteome Software Inc). Identified peptides were accepted with probabilities >95% (Scaffold local false discovery rate algorithm). Accepted protein identifications had a probability >99.0% and contained at least one identified peptide.

### Negative stain sample preparation of hSAGA, data collection and processing.

Here, 400 mesh Cu grids were cleaned three times (in ethanol, water, ethanol) by sonication for 5 min and dried on filter paper. A petri dish was filled with water forming a meniscus and wiped off with lens paper. One drop of 1% (w/v) nitrocellulose in amyl acetate was added to the surface, forming a thin film. Cleaned grids were deposited on the film with the polished side facing down. The grids were transferred with parafilm onto filter paper with the nitrocellulose facing up and dried overnight. Grids were coated with carbon by evaporation using an Edwards Auto306 (10<sup>-6</sup> mbar, 6 A, 6 s). Before sample adsorption, grids were glow discharged (30 s, 15 W, Tergo EM PIE scientific). Human SAGA was diluted (2x) in dilution buffer (25 mM HEPES pH 7.5, 0.2 mM EDTA, 6 mM MgCl<sub>2</sub>, 0.2 M NaCl, 3% (w/v) D(+)-Trehalose), 3 µl were applied to a grid and adsorbed for 1 min. The grid was washed and stained, respectively, by swirling it five times on a 50 µl drop of 2% (w/v) uranyl formate for 10 s (each), blotted and dried in an air stream.

Data was collected on a Tecnai F20 (Thermo Fisher Scientific), using Legion<sup>44</sup> (Fig. 2 and Table 1). Micrographs were contrast transfer function (CTF) corrected using CTFind v.4.1.13 (ref.<sup>45</sup>). Particles were picked using Gaussian LoG picker in Relion-3.1 (ref.<sup>46</sup>), extracted with a box size of 300 × 300 pixels and subjected to reference-free two-dimensional (2D) classification. Particles from the best classes (32%) were used for initial model generation using the statistical gradient descent method<sup>47</sup> in Relion-3.1 (ref.<sup>46</sup>). Particles were classified by a series of three-dimensional (3D) and 2D classifications with and without alignment (Extended Data Fig. 2a). Classification revealed one class without the (ordered) TAF6L HEAT domain and SPL module. Particles with and without this region were separated by multi-reference 3D classification. The best reconstruction was refined and classified again by alignment-free 3D classification. Combined classes that yielded the highest resolution were refined and then postprocessed in Relion v.3.1 (ref.<sup>46</sup>) (Extended Data Fig. 2b).

### Cryo-EM sample preparation of hSAGA, data collection and processing.

Quantifoil Au 300 mesh UltrAuFoil R1.2/1.3 polyethylenimine (PEI)/graphene-oxide grids were prepared according to established protocols<sup>48</sup>. The grids were used for freezing within 2–4 h.

All grid preparation steps were done on ice. Here, 3 µl of undiluted hSAGA was transferred into a 0.5 ml non-stick tube and crosslinked by mixing with 0.6 µl of crosslinking buffer (25 mM HEPES pH 7.8, 0.2 M NaCl, 0.2 mM EDTA, 0.5 mM TCEP, 0.01% (v/v) NP40, 10% (v/v) glycerol, 6 mM bis(sulfosuccinimidyl)suberate) and incubated for 5 min. A graphene-oxide grid was picked up with Vitrobot tweezers, the sample was transferred to the grid and incubated for 2 min in a saturated humidity chamber. Afterward, the grid was washed five times by submerging and swirling for 5 s (each) in 230 µl of wash buffer (25 mM HEPES pH 7.8, 0.2 M NaCl, 0.2 mM EDTA, 0.5 mM TCEP, 0.01% (v/v) NP40, 2.5% (v/v) glycerol) in a five-well Teflon block. Without letting the grid dry, excess solution was blotted off at a 90° angle and 4 µl of wash buffer were added immediately. The grid and tweezers were mounted into a Vitrobot Mark IV (Thermo Fisher Scientific), blotted with fresh filter paper (blot force 0, 3 s) and plunge frozen into liquid ethane.

Data were collected with SerialEM<sup>49</sup> and 3 × 3 multishot acquisition on a Titan Krios G2 (Thermo Fisher Scientific) (Table 1). Videos were whole-frame motion corrected and binned (2×) using the Relion-3.1 (ref. 46), CTF corrected using CTFind v4.1.13 (ref. 45) and sorted manually. Particles were picked using the Gaussian LoG picker in Relion-3.1 (ref. 46) and extracted with 8× binning (Extended Data Fig. 3a) and a box size of 45 × 45 pixels. Graphene-oxide edges were removed by 2D classification before hSAGA particles could be classified. After removing most graphene-oxide edges, particles were reextracted with recentering (4× binned, 90 × 90 box size) and reclassified in 2D. The negative stain reconstruction was low-pass filtered to 50 Å and used as reference model for initial 3D classification. Each class was subclassified by alignment-free 2D classification to remove particles close or on graphene-oxide edges. The remaining particles were subjected to 3D classification, recentered in the box by applying a coordinate transformation to the particle alignment parameters using a custom python script, reextracted with recentering, without binning and placed in a box size of 360 × 360 pixels, then subjected to a consensus 3D refinement. Afterward, the particles were subjected to two rounds of Bayesian polishing, 3D refinement, CTF refinement and alignment-free 3D classification (tau = 20) (Extended Data Fig. 3a). A final round of 3D classification, refinement and postprocessing yielded a reconstruction at 2.93 Å (Extended Data Fig. 3b–d). High variability and low local resolution were observed at the TRRAP N terminus and the HEAT repeat cradle in close proximity as well as around the surface of the core module. Low-pass filtering and B factor blurring slightly improved the interpretability of the map in these regions. Further improvement was made by multibody refinement (Extended Data Figs. 3e and 8) of the core and TRRAP modules, although the resolution did not improve. Various density modification and map enhancement methods were tested, and the greatest improvements in variable and surface exposed regions were obtained by applying the spiral phase transform in LocSpiral<sup>50</sup> to all reconstructions. This process revealed additional peptide connections and density fragments of the disordered TAF6L HEAT domain (Extended Data Fig. 3e,i,j). A principal component analysis of the multibody refinement showed a high degree of flexibility between the core and TRRAP modules (Extended Data Fig. 8a), which can also be observed by 3D variability analysis in Cryosparc v2.15.0 (refs. 47,51). Signal subtraction after recentering and reextraction was attempted to detect density for the HAT and DUB modules but was not successful, presumably due to a high degree of conformational as well as potential compositional heterogeneity (metazoan SAGA has also been observed to occur without these modules<sup>52,53</sup>). Nevertheless, early samples that had been stabilized by GraFix<sup>54</sup> revealed some 2D negative stain classes with highly variable density that is consistent with the suggested locations based on comparisons with ySAGA and the position of the N-terminal end of ATNX7 (Extended Data Fig. 4g). Masking and map transformations were carried out using UCSF Chimera<sup>55</sup> and Relion-3.1 (ref. 46). All resolution estimates are based on the 0.143 threshold criterion of the gold-standard Fourier shell correlation (FSC)<sup>56</sup> of two independently refined half sets in Relion-3.1 (ref. 46), after accounting for correlations introduced by masking<sup>57</sup>. Local resolution was estimated using Relion-3.1 (ref. 46).

**Cryo-EM sample preparation of hSAGA with TBP, data collection and processing.** Human SAGA was mixed with a sixfold molar excess of human full length TBP and incubated for 5.2 h on ice. Grids were frozen and data were collected and processed in the same way as described above, but no additional density corresponding to TBP could be observed.

**Modeling and refinement.** For model building in Coot<sup>58</sup> maps were converted to structure factors using phenix.map\_to\_structure\_factors<sup>59</sup>, allowing low-pass filtering and variable B-sharpening or blurring in Coot<sup>58</sup>. Models were built into the postprocessed, multibody-refined and LocSpiral<sup>50</sup> filtered maps (Extended Data Fig. 3e–g). A fragmented initial model of secondary structure elements in TRRAP was generated using phenix.map\_to\_model<sup>59</sup>, manually corrected and completed in Coot<sup>58</sup>. A homology model of the TAF5L WD40 propeller was generated using SwissModel<sup>60</sup> (based on Protein Data Bank (PDB) accession code 6F3T, <sup>61</sup>) and rigid-body fitted in Coot<sup>58</sup>. The remaining model was built de novo in Coot<sup>58</sup>, guided by homology models based on human TFIID<sup>9</sup> and ySAGA<sup>4,5</sup>. Regions with low confidence in register assignment were modeled as poly-alanines (assigned as unknown, UNKs). Before real-space refinement in Phenix, atomic B factors were reset to 90 Å<sup>2</sup>, the model was protonated using phenix.ready\_set<sup>59</sup> and sanity checked as well as geometry minimized using jelly<sup>62</sup> (GlobalPhasing). Afterward, the model was refined using Rosetta<sup>63</sup>, validated using phenix.molprobity<sup>59</sup> and optimized in Coot<sup>58</sup>. Secondary structure restraints were generated using phenix.secondary\_structure\_restraints and corrected after manual inspection. A final refinement was carried out with phenix.real\_space\_refine<sup>59</sup> (1.18–3861) against the complete LocSpiral<sup>50</sup> filtered map using default parameters plus secondary structure restraints, rotamers.fit=outliers\_and\_poormap and rotamers.tuneup=outliers\_and\_poormap (Extended Data Fig. 3g). Model statistics were calculated using phenix.molprobity<sup>59</sup> (Table 1). Refinement against the regular postprocessed map resulted in almost identical statistics with an all-atom r.m.s.d. of 0.400 Å. All maps used for model building and refinement were deposited in the Electron Microscopy Data Bank (EMDB). Map versus model FSC was calculated using phenix.mtriage<sup>59</sup> (Extended Data Fig. 3f). The InSp<sub>6</sub> ligand was identified

by density fit and homology to mTORC2 (ref. 31) and one out of two possible conformations was modeled (Fig. 3f and Extended Data Fig. 6e). In analysis of our cryo-EM structure, separating the core and TRRAP modules improved map quality and revealed additional features on surface exposed regions after LocSpiral filtering<sup>50</sup> (Extended Data Fig. 3i,j). In particular, analysis of the region where SUPT3H, SUPT7L, TADA1 and TRRAP meet suggested alternative main chain conformations that could not be sorted out by classification. The highly variable region between SUPT3H, TADA1 and TRRAP corresponds to the approximate position where TBP binds to ySAGA (Extended Data Fig. 8b,c). The presence of multiple subunit isoforms, identified by mass spectrometry (Supplementary Table 2), did not affect modeling. Differences of isoforms are primarily located in disordered regions and were addressed according to PDB standards with remarks.

A model for the negative stain reconstruction was generated by rigid-body fitting without coordinate refinement in phenix.real\_space\_refine<sup>59</sup> using the protein part of the cryo-EM model, a homology model of the TAF6L HEAT domain (generated with SwissModel<sup>60</sup> and based on human TAF6, ref. 9, PDB 6MZL), and SF3B3/SF3B5 from the SF3b core complex<sup>10</sup> (PDB 5IFE) (Table 1 and Extended Data Fig. 2d–i). Before fitting, expression tags in SF3B3 were deleted and the TAF6L HEAT domain was mutated to poly-alanines (annotated as UNKs) to reflect the absence of an authentic high- or medium-resolution structure for this region.

**Structural analysis and visualization.** Coordinate transformations and manipulations were carried out using CCP4 tools<sup>64</sup>. Structures were compared using PDBeFold<sup>65</sup> and interfaces were analyzed using QtPISA v2.1.0 (ref. 66). Relative angles between variable regions/domains (for example, TAF5(L) NTDs) of related structures with a common reference domain (for example, TAF5L WD40) were calculated by prealigning all structures to the reference domain of hSAGA using secondary structure matching. The center of masses of the hSAGA reference domain (for example, TAF5L WD40), the hSAGA variable domain (for example, TAF5L NTD or TRRAP ΨPIKK) and the hSAGA variable domain after superposition on the corresponding domain in related structures using secondary structure matching (for example, ySAGA Taf5 NTD or Tra1 ΨPIKK) were calculated. Center of masses were calculated in PyMOL (The PyMOL Molecular Graphics System, v2.4.0 Schrödinger, LLC.) and angles between corresponding vectors were calculated using python. Related structures were identified using PDBeFold (70% query/70% target)<sup>65</sup>. Structure figures were generated using PyMOL, ChimeraX (UCSF, 2020-01-10) and Adobe Illustrator. Electrostatic surfaces were generated using the APBS<sup>66</sup> plugin in PyMOL. Videos were generated using ChimeraX<sup>67</sup> (UCSF, 2020-01-10), Adobe Premier and ffmpeg (<https://ffmpeg.org>). Plots were generated using python. Reported contour levels for maps are defined as  $\sigma = \text{density threshold}/r.m.s.$

**Sequence analysis.** In total, 23 metazoan homologs of hSAGA with a complete set of all 20 subunits were retrieved from databases. Sequence alignments were generated using the Clustal Omega<sup>68</sup> executable in Geneious Prime v2021.0.3. Sequence conservation figures were generated by aligning all subunit sequences of all 23 metazoan SAGAs with the sequences of the molecular model of hSAGA. Alignments were combined and conservation scores were calculated using AL2CO<sup>69</sup> and used for coloring in PyMOL.

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

## Data availability

Cryo-EM maps and refined coordinates were deposited in the EMDB with accession codes EMD-23027 and EMD-23028 and in the PDB with accession codes 7KTR and 7KTS. The cell line can be provided on request. Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to enogales@lbl.gov. Source data are provided with this paper.

## Code availability

Custom computer code is available on Github (coord\_transform\_to\_star, [https://github.com/dominikaherbst/cryo-em\\_scripts](https://github.com/dominikaherbst/cryo-em_scripts)).

## References

- Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, <https://doi.org/10.7554/eLife.25776> (2017).
- Herrera, F. J., Yamaguchi, T., Roelink, H. & Tjian, R. Core promoter factor TAF9B regulates neuronal gene expression. *eLife* **3**, e02559 (2014).
- Suloway, C. et al. Automated molecular microscopy: the new Legion system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, <https://doi.org/10.7554/eLife.42166> (2018).

47. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
48. Patel, A., Toso, D., Litvak, A. & Nogales, E. Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.08.434344> (2021).
49. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastrorade, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
50. Kaur, S. et al. Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nat. Commun.* **12**, 1240 (2021).
51. Punjani, A. & Fleet, D. J. 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
52. Soffers, J. H. M. et al. Characterization of a metazoan ADA acetyltransferase complex. *Nucleic Acids Res.* **47**, 3383–3394 (2019).
53. Li, X. et al. Enzymatic modules of the SAGA chromatin-modifying complex play distinct roles in *Drosophila* gene expression and development. *Genes Dev.* **31**, 1588–1600 (2017).
54. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
55. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
56. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
57. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **60**, 2126–2132 (2004).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
60. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
61. Antonova, S. V. et al. Chaperonin CCT checkpoint function in basal transcription factor TFIID assembly. *Nat. Struct. Mol. Biol.* **25**, 1119–1127 (2018).
62. Bricogne, G. et al. BUSTER v.2.10.3 (Global Phasing Ltd, 2011).
63. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
64. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D.* **50**, 760–763 (1994).
65. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2256–2268 (2004).
66. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
67. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* <https://doi.org/10.1002/pro.3943> (2020).
68. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
69. Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700–712 (2001).

### Acknowledgements

We thank J. Fang for help with hSAGA purification, A.B. Patel for discussion and the graphene-oxide protocol, P. Grob and D. Toso, J. Remis and P. Tobias in the Cal-Cryo facility at UC Berkeley for microscope access and support, A. Chintangal for computing support, E. Spooner and the Whitehead mass spectrometry facility for mass spectrometry analysis, S. Zheng for help with HeLa cell culture, E. Borbon for help with HeLa cell collection and C. He for help with cloning of constructs and cell maintenance. This work was funded by the National Institute of General Medical Sciences grant nos. R01-GM63072 and R35-GM127018 to E.N.D.A.H. was supported by EMBO ALTF 1002-2018 and SNSF P2BSP3\_181878, and M.N.E. by National Institutes of Health training grant no. T32GM098218. E.N. and R.T. are Howard Hughes Medical Institute Investigators.

### Author contributions

E.N., R.T. and X.D. supervised the work. R.K.L. piloted cell culture work, purifications, initial negative stain EM and generated the cell line together with C.D.-D. G.M.D. cloned CRISPR plasmids and performed genotyping of the edited cell line. M.N.E. performed large scale cell culture and nuclei harvesting, purified hSAGA, performed biochemical validation experiments and analyzed the mass spectrometry data. D.A.H. carried out EM sample preparation, EM, data processing, model building and refinement, structural analysis and sequence analysis. Q.F. processed the TBP data set. D.A.H., M.N.E. and E.N. wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

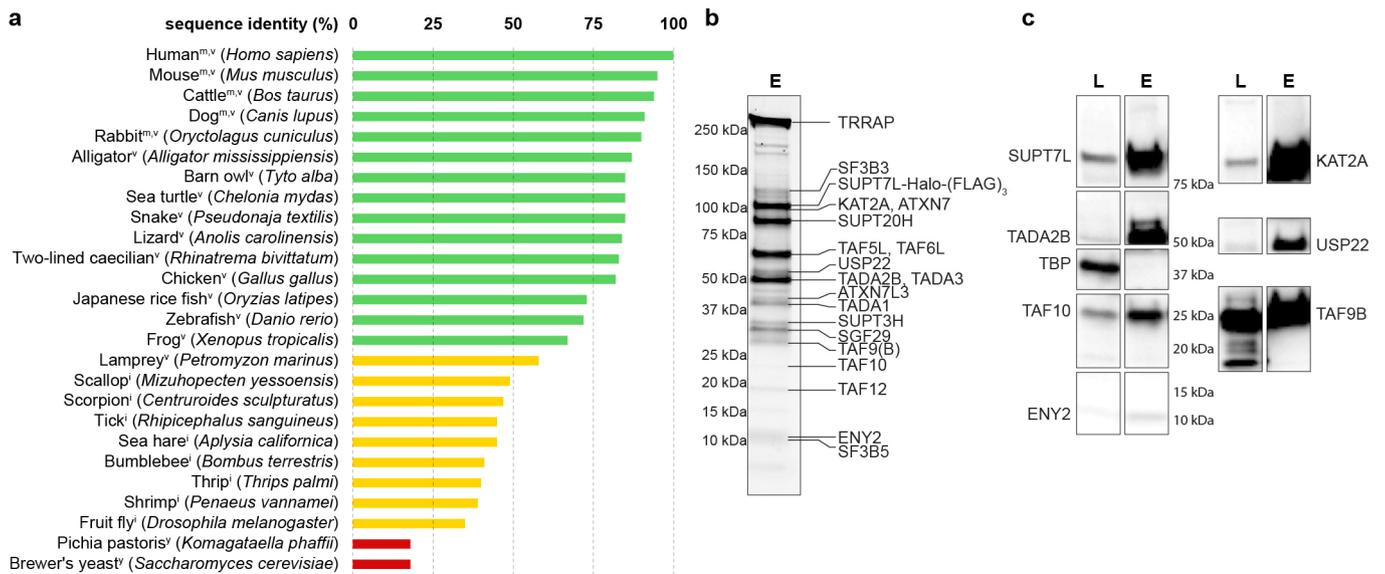
**Extended data** are available for this paper at <https://doi.org/10.1038/s41594-021-00682-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41594-021-00682-7>.

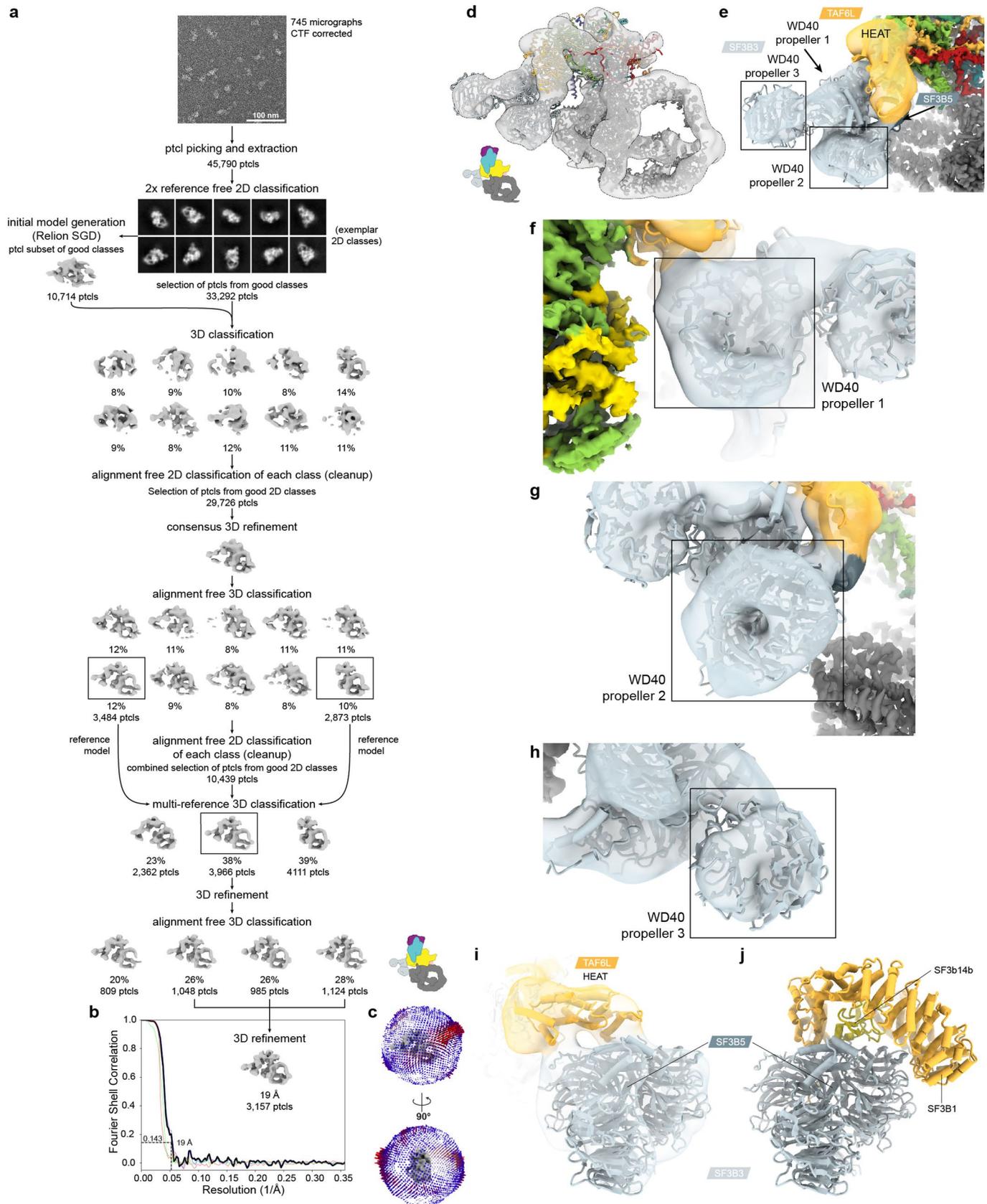
**Correspondence and requests for materials** should be addressed to Eva Nogales.

**Peer review information** *Nature Structural and Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Beth Moorefield was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

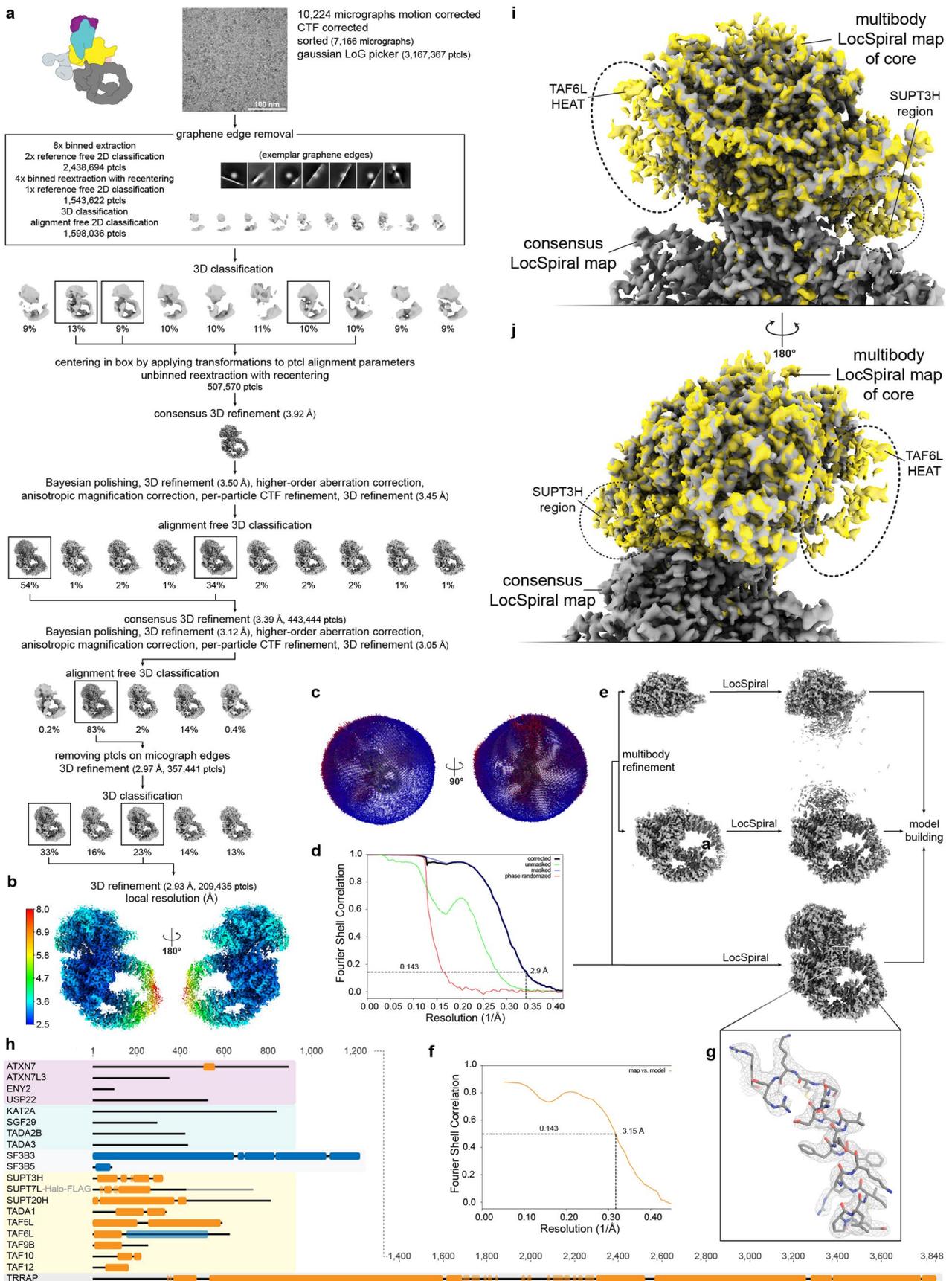


**Extended Data Fig. 1 | Conservation and Purification of hSAGA.** **a**, Average sequence conservation of SAGA subunits in metazoan and yeast (relates to Supplementary Table 1). Indices indicate classification into mammals (m), vertebrate (v, green), invertebrate (i, yellow), and yeast (y, red). **b**, 4-20% gradient SDS-PAGE gel stained with Flamingo fluorescent protein stain (BioRad) of the hSAGA FLAG elution (E) with subunits labeled based on their predicted molecular weight. **c**, Western blot probing for DUB, HAT, and core subunits to verify the presence of these modules in the sample used for grid preparation. TBP did not co-purify with hSAGA. The lysate lane (L) corresponds to 0.0004% of the total input and the right lane (hSAGA Elution, E) corresponds to 4.65% of the final elution. Blots were cropped. Experiments in b and c were repeated twice with similar results.



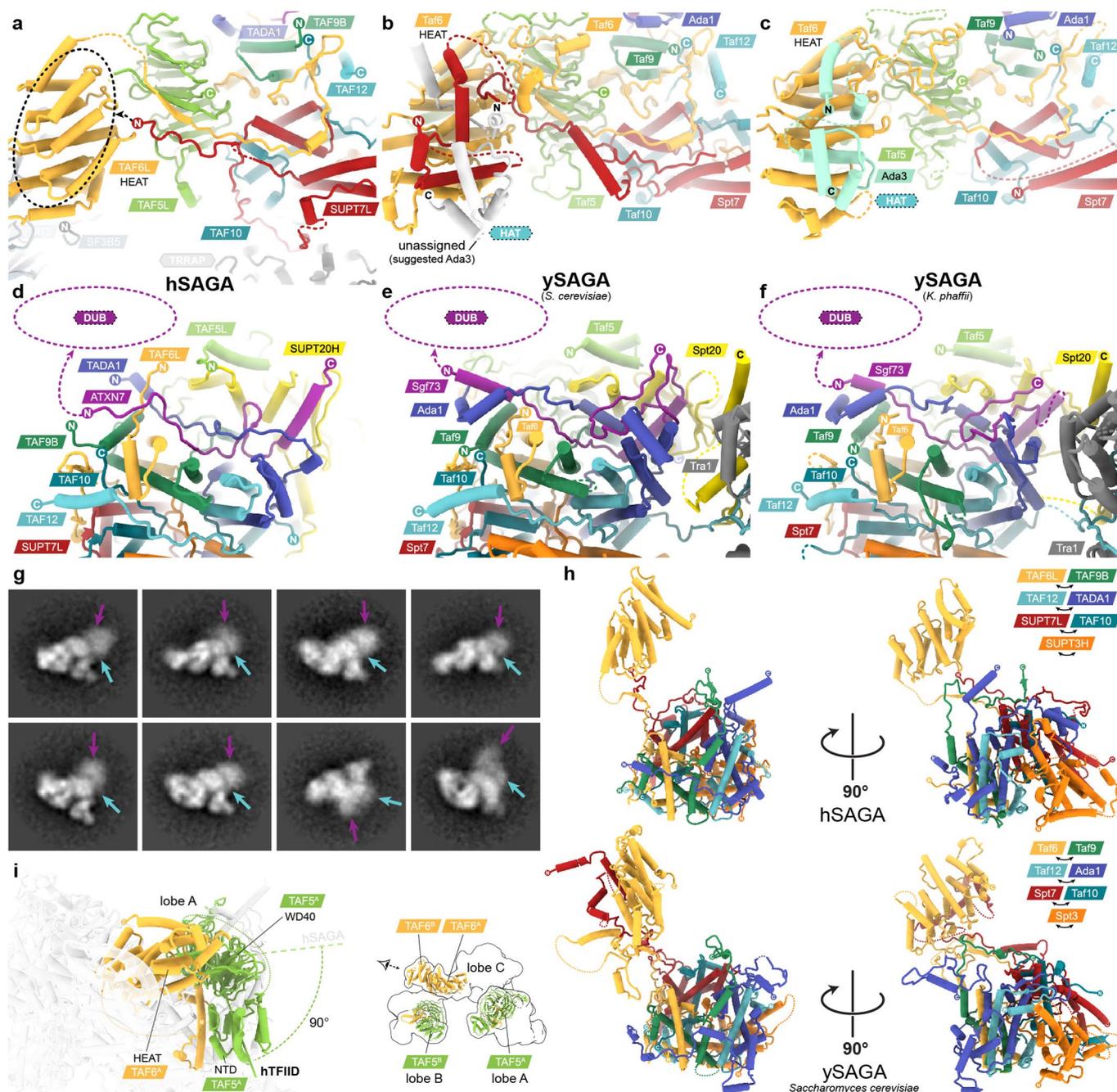
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Negative stain processing scheme and model fit.** A representative section of one of 745 micrographs is shown. After initial 2D classification, particles (ptcls) from the best classes were used for initial model generation. The data was cleaned up by 3D classification followed by alignment-free 2D classification. Particles from all good classes were subjected to a consensus 3D refinement followed by alignment-free 3D classification. All except for one class revealed fuzzy density for the TAF6L HEAT and SPL region. Subsequently, all classes were cleaned up individually by alignment-free 2D classification and combined in a multi-reference classification using the two best models with and without TAF6L HEAT and SPL region. The best class including this region was subjected to 3D refinement, alignment-free 3D classification, and to a final refinement using particles of the class combination that yielded highest resolution. **b.** Final map and FSC plot. **c.** Angular distribution. **d.** Final map (contoured at  $4.9 \sigma$ ) and rigid body fit of SF3B3/SF3B5 (from PDB: 5IFE), a homology model of the TAF6L HEAT domain, and the cryo-EM structure from this study. **e-h.** Close-up view on the SPL module region of the hybrid map shown in Fig. 1c. The rigid-body fit is shown with translucent map surfaces. All domains fit precisely in the negative stain density, which shows clear central holes for the three WD40 propellers (f-h) of SF3B3. **i, j.** Comparison of the SF3B3/SF3B5 integration in hSAGA and the SF3b complex. **i.** The SPL module (SF3B3 and SF3B5 subunits) binds to the concave surface of the TAF6L HEAT domain. The negative stain map of hSAGA is shown in translucent white (contoured at  $4.8 \sigma$ ). **j.** Crystal structure of the SF3b complex<sup>10</sup> (PDB: 5IFE). The TAF6L HEAT domain of hSAGA is replaced by the SF3B1 HEAT repeat domain in the SF3b complex. Both domains share an overlapping binding region on the SF3B5 surface.

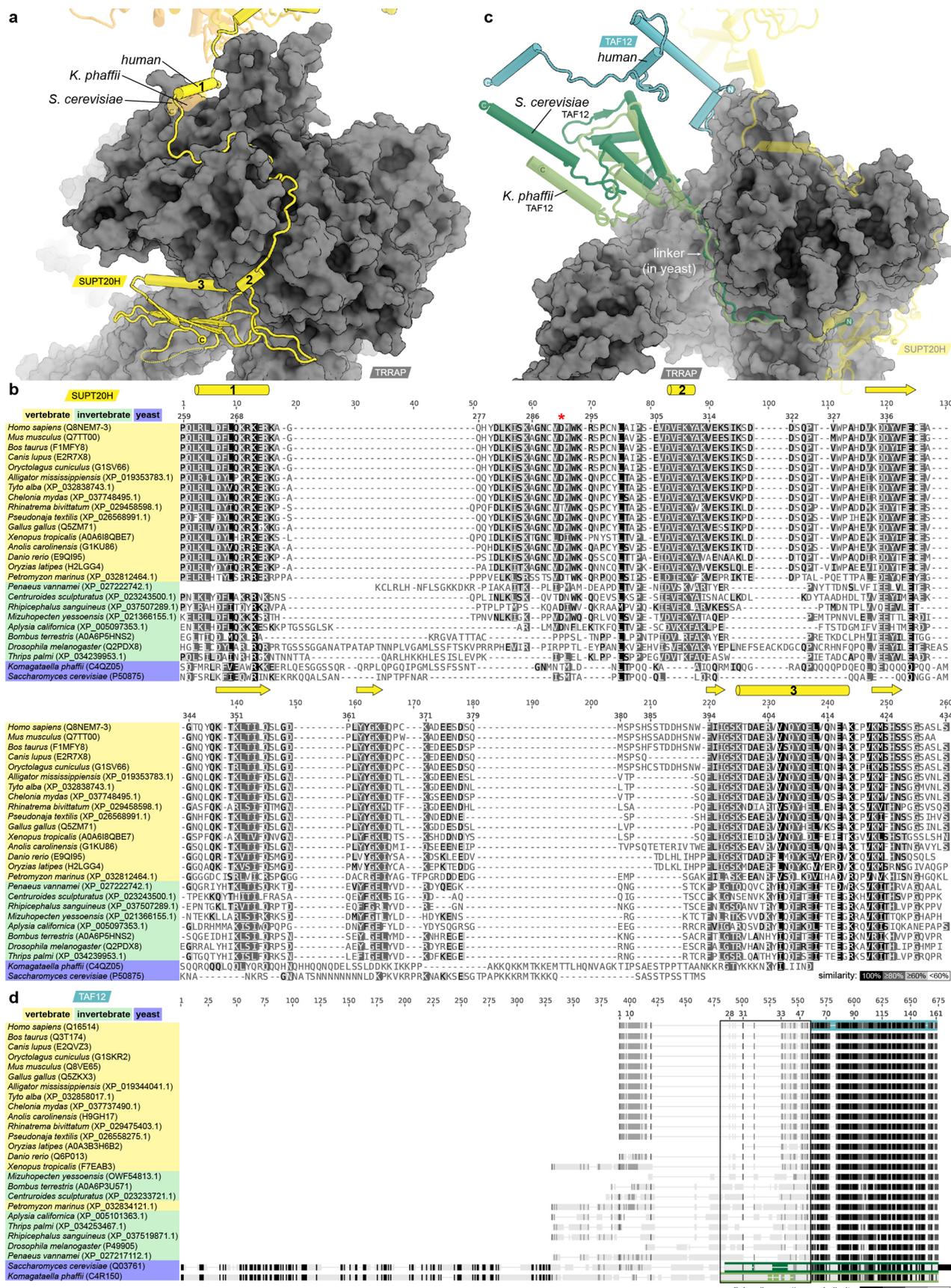


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Cryo-EM processing and model building.** A representative section of one of 10,224 micrographs is shown. Graphene oxide (GO) edges were removed in cycles of initial 2D and one 3D classification. The negative stain reconstruction (see Extended Data Fig. 2) was used as initial model. 3D classes were centered in the box by applying a coordinate transformation to the alignment parameters, and unbinned particles were re-extracted with recentering. Particles were filtered for high-resolution features in cycles of 3D refinement, classification, Bayesian polishing, and CTF refinement as indicated. **b.** Postprocessed map (B-sharpened with  $-51.1 \text{ \AA}^2$ , contoured at  $4.9 \sigma$ ) with local resolution. **c.** angular distribution. **d.** Fourier Shell Correlation (FSC). **e.** Multibody refinement improved map quality, but not the overall resolution. Considerable improvement of map quality was achieved by filtering with LocSpiral<sup>50</sup>. The model for the core and TRRAP was built into the LocSpiral filtered maps of the multibody refinement. The interface between these regions was built using the full map and used for model refinement. Refinement against the postprocessed map (**b**) resulted in the same model, with virtually identical statistics and an all-atom r.m.s.d. of  $0.400 \text{ \AA}$ . Maps are contoured at (regular/LocSpiral): Core  $11.2 \sigma/9.2 \sigma$ , TRRAP  $7.9 \sigma/9.0 \sigma$ , full  $6.9 \sigma$ . **f.** Map vs. model FSC using the postprocessed map shown in **b**. **g.** The refined map shows well defined secondary structure elements and side chains (contoured at  $9.0 \sigma$ ). **h.** Model-sequence coverage. Sequences of all subunits are indicated as horizontal lines (black) and modeled regions as overlaying boxes (orange: visualized by cryo-EM; blue: visualized only by negative stain; translucent: regions with unclear register assignment (unknown, UNKs)). **i, j.** The LocSpiral filtered multibody map of the core reveals additional density corresponding to the poorly ordered TAF6L HEAT domain, and to SUPT3H in the cleft between the core and TRRAP module (both contoured at  $5.9 \sigma$ ).

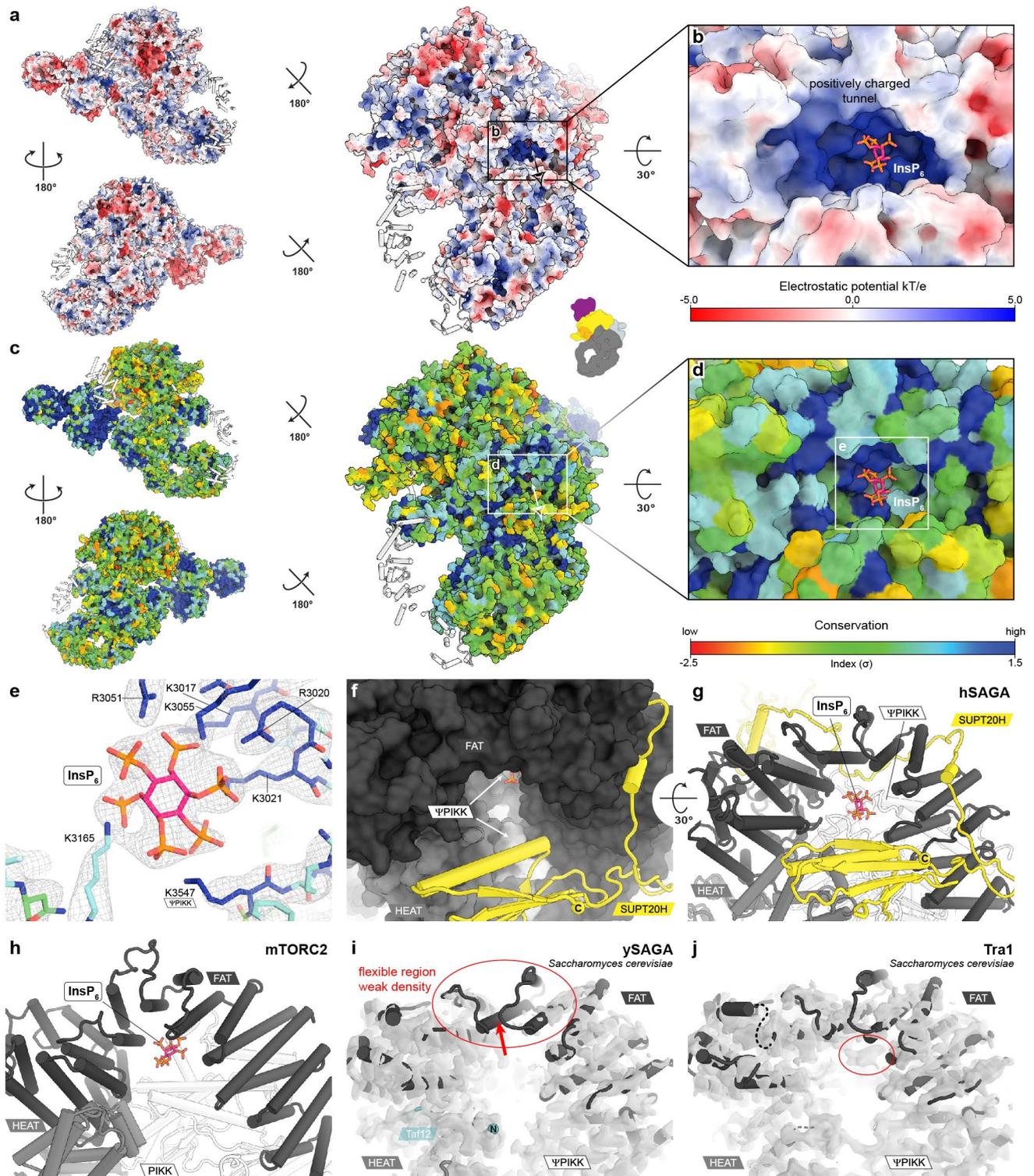


**Extended Data Fig. 4 | The core module and tethering of its peripheral modules in human and yeast.** **a-c.** Tethering of the HAT module: **a.** The N-terminus of SUPT7L runs parallel to the TAF6L linker, which connects to the HEAT domain, along the surface of the core and ends with its N-terminus in close proximity to the HEAT domain. **b.** In *Saccharomyces cerevisiae*<sup>4</sup> (PDB: 6T9I), the Spt7 linker further extends towards the convex surface of the TAF6L HEAT domain and interacts with an unassigned region. **c.** The same region in *Komagataella phaffii*<sup>5</sup> was assigned as the Ada3 subunit of the HAT module (PDB: 6TBM). The similar location of the SUPT7L N-terminus suggests a similar interaction and connectivity for the HAT module in hSAGA. **d-f.** Tethering the DUB module. **d.** The ATXN7 subunit of the core and the DUB module (Sgf73 in yeast) is similarly integrated into the core module as in ySAGA (**e, f**), suggesting a similar relative attachment of the human DUB. **g.** GraFix<sup>54</sup> crosslinked negative stain class averages revealed extra density at the anticipated locations for the HAT (cyan arrow) and DUB (purple arrow) modules. **h.** The core of hSAGA and ySAGA (PDB: 6T9K) builds on common HF elements. Only HF containing subunits are shown. Architectural differences are created by local variations outside of the HFs. HF dimerization is indicated by arrows below the subunit labels. **i.** Comparison of TAF5 and TAF6 architecture within lobe A of human TFIID (canonical state<sup>9</sup>, PDB: 6MZL). TFIID contains two copies of TAF5 and TAF6, with one TAF5 located in lobe A (TAF5<sup>A</sup>) and the other one in lobe B (TAF5<sup>B</sup>), and with the two TAF6 HEAT domains (TAF6<sup>A</sup>, TAF6<sup>B</sup>) in lobe C (shown on the right). Compared to hSAGA, the TFIID TAF6 HEAT domains are differently arranged relative to TAF5, and they act to bridge lobes B and C. The TFIID TAF5 NTD is rotated by 90°, leading to a divergent architecture.



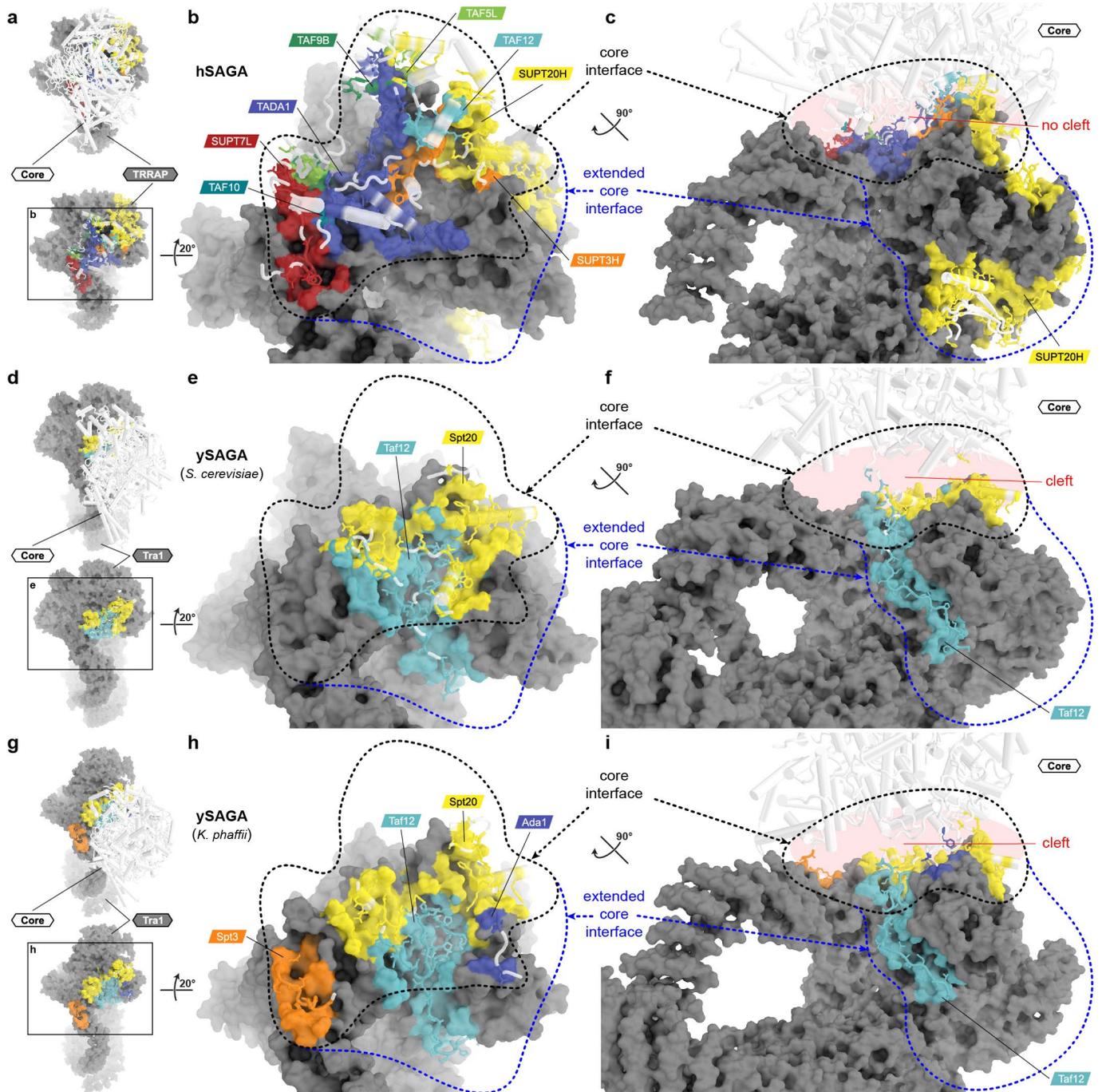
Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Human versus yeast interactions between TRRAP/Tra1, SUPT20/Spt20 and TAF12/Taf12.** **a.** Location of the SUPT20H/Spt20 C-terminal region after superposition of human TRRAP and yeast Tra1. The C-terminal helix of yeast Spt20 aligns with helix one of the SUPT20H linker. **b.** The sequence alignment of the SUPT20H/Spt20 C-terminal regions for 24 metazoan (SUPT20H) and two yeast (Spt20) species shows that the SUPT20H CTD is highly conserved in vertebrates, while it does not appear to exist in yeast. Secondary structure elements are indicated above the alignment. \*: D291 forms a salt bridge with TRRAP R3746 (Fig. 5f). Vertebrate and invertebrate sequences were pre-aligned to human SUPT20H, regions corresponding to the structured part in **a** were extracted and realigned with the yeast sequences corresponding to the region from helix 1 in **a** to the C-terminus. **c.** Relative location of the TAF12 N-terminal region, based on the superposition shown in **a**. In yeast, an N-terminal linker of Taf12 wraps around the inside of the Tra1 FAT domain, while human TAF12 contacts TRRAP in a different location. The structured N-terminus of yeast Taf12 is located in the same relative position as the human SUPT20H CTD. **d.** Zoomed-out sequence alignment of 24 metazoan and two yeast TAF12/Taf12 subunits. Structured regions are colored as in **c** and the region corresponding to the linker in yeast is indicated. Aligned regions in **b** and **d** are colored by similarity in gray scale (annotated in **d**). In yeast, Taf12 contains a considerably longer N-terminus that appears to be unique to yeast. Sequences are labeled as: Scientific organism name (Uniprot or NCBI accession code). The organism selection corresponds to Extended Data Fig. 1a and Supplementary Table 1.

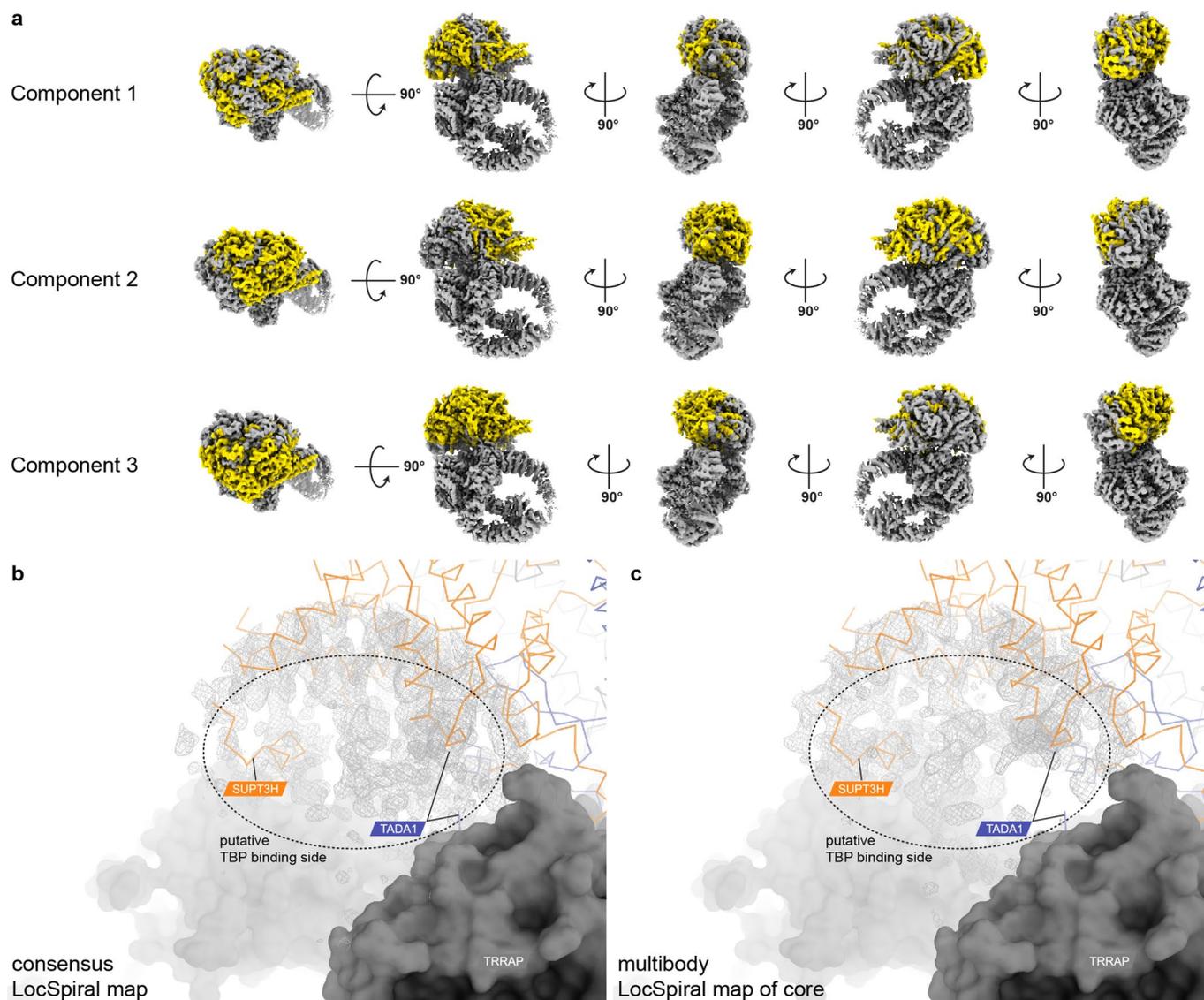


Extended Data Fig. 6 | See next page for caption.

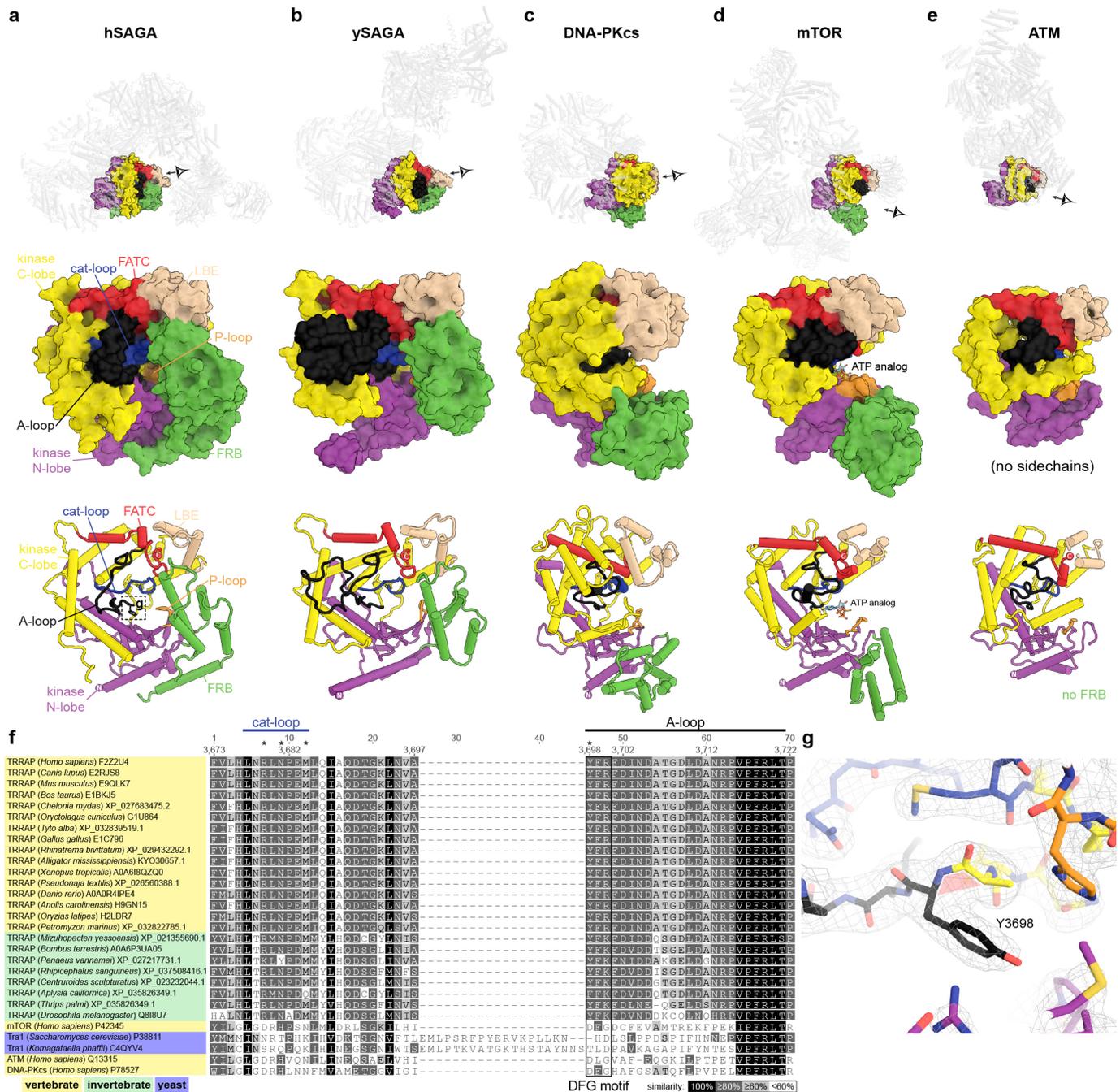
**Extended Data Fig. 6 | Electrostatic surface, conservation, and binding of inositol hexakisphosphate (InsP<sub>6</sub>) to TRRAP.** **a.** Electrostatic surface representation of hSAGA shown from three different views. Only regions with all-atom models have been included. Regions of lower sequence assignment confidence (unknown, UNKS) were excluded from the calculation and are shown as white cartoon. A highly positive charged tunnel between the FAT, HEAT, and ΨPIKK domains of TRRAP and SUPT20H is indicated. **b.** Close-up view of the InsP<sub>6</sub> binding pocket within this tunnel. **c.** Same views as in **a** colored by sequence conservation (white cartoon as in **a**). **d.** Close-up view of the InsP<sub>6</sub> binding pocket. InsP<sub>6</sub> (shown in stick representation) is bound by a half-ring of highly conserved residues. **e.** Close-up view of the atomic model and the LocSpiral filtered multibody map (contoured at 11  $\sigma$ ) of TRRAP. The view corresponds to a back view of Fig. 3f (rotated 180°). Residues involved in InsP<sub>6</sub> binding are indicated. All labeled residues are part of the TRRAP FAT domain except for K3547 (within ΨPIKK). Atoms are colored by conservation (carbon, see panel d), pink (carbon of InsP<sub>6</sub>), red (oxygen), blue (nitrogen), or orange (phosphorus). **f.** View as indicated by boxes in a and c (colored by domains and subunits). **g, h.** Comparison of InsP<sub>6</sub> binding in hSAGA and in human mTORC2<sup>31</sup> shown in cartoon representation. In both complexes InsP<sub>6</sub> binds in a similar location between the FAT and pseudo- and kinase domains of hSAGA and mTORC2, respectively. **i.** In the  $\gamma$ SAGA complex structures (here PDB: 6T9J) the central part of the FAT domain is poorly resolved (red circle) and presumably flexible (translucent map contoured at 11  $\sigma$ ). This region contains residues (red arrow) involved in InsP<sub>6</sub> binding in hSAGA (for example R3051 and K3055, see panel e). **j.** Unattributed density (red circle) in the earlier determined structure of isolated Tra1 (PDB: 5OJS, translucent map contoured at 7.5  $\sigma$ ).



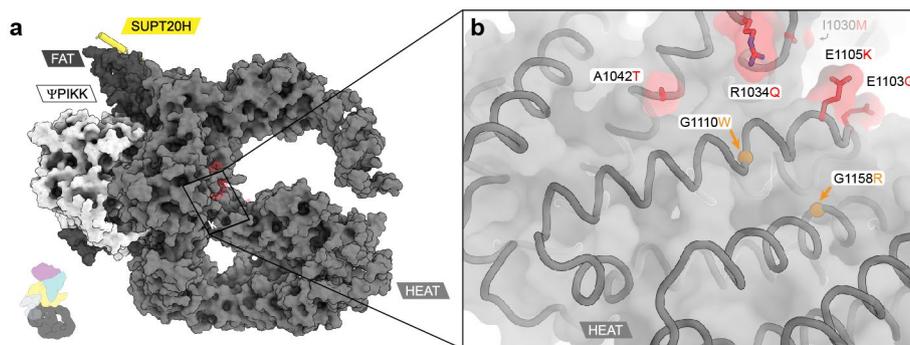
**Extended Data Fig. 7 | Comparison of the core-TRRAP/Tra1 interface in human and yeast.** **a.** Top view on the hSAGA core (white cartoon) and TRRAP module (gray surface). Parts of the core not in direct proximity to the interface have been removed in the lower panel. Interfacing residues on the TRRAP surface are colored based on their closest core subunit. Interfacing residues of core subunits are shown in stick representation and colored by subunit. **b.** Magnified region from the box in **a**. The interface corresponding to the footprint of the core on the TRRAP surface is indicated with a black dashed outline (core interface). The interface created by extensions of core subunits that wrap around the TRRAP is indicated with a blue dashed outline (extended core interface). In hSAGA the latch and SUPT20H CTD (residues 274-428) contribute to the extended core interface. **c.** Side view of the core and extended interface with the complete core shown as white cartoon. In contrast to ySAGA (**f, i**), hSAGA has no cleft between the core and TRRAP modules. **d, f.** Same view as in **a-c** for *S. cerevisiae* ySAGA (PDB: 6T91)<sup>4</sup>. **g, h.** Same view as in **a-c** for *K. phaffii* ySAGA (PDB: 6TB4)<sup>5</sup>. The colored surface regions on TRRAP/Tra1 correspond to the colors of the contacting core subunits and thus show the different participation of core subunits to the interface. **f, i.** In both ySAGA structures the core and Tra1 modules are separated by a cleft.



**Extended Data Fig. 8 | Conformational variability in hSAGA between the core and TRRAP domains. a.** Principal component analysis of the multibody refinement reveals several tilting and swiveling motions between the core and TRRAP module. **b, c.** The cleft between SUPT3H, TADA1 (colored ribbon representation), and TRRAP (surface representation) reveals highly variable density, shown in grey mesh with a radius of 20 Å, in the LocSpiral filtered consensus maps of the core (b, contoured at  $6\sigma$ ) and the multibody (c, contoured at  $9.2\sigma$ ). Other core subunits are indicated in light grey ribbon. A region of high variability is shown with the dashed oval (Relates to white oval in Fig. 4b).



**Extended Data Fig. 9 | TRRAP ΨPIKK comparison and its integration into the HEAT repeat and FAT domain scaffold.** **a-e.** ΨPIKK and its integration in hSAGA (**a**) and ySAGA<sup>4</sup> (**b**, PDB: 6T9I), compared with the PIKK domains in DNA-PKcs<sup>70</sup> (**c**, PDB: 6ZFP), mTOR<sup>71</sup> (**d**, PDB: 6BCX), and ATM<sup>72</sup> (**e**, PDB: 5NPO). The panels below the overview show a close-up view as surface and cartoon representation of the active site entrance, as indicated in the top panels. Kinase elements are colored as indicated for hSAGA (FRB: FKBP-Rapamycin binding, P-loop: phosphate binding loop, LBE: mLST8 binding element, cat-loop: catalytic loop, FATC: FRAP-ATM-TRRAP C-terminus, A-loop: activation loop). For mTOR the kinase crystal structure with ATPyS<sup>73</sup> (PDB: 4JSP) is shown. In agreement with the comparison by Díaz-Santín *et al.*<sup>17</sup>, the widely opened active site entrance in active kinases (**c-e**) is narrowed to a small tunnel in the pseudo-kinase in TRRAP (**a**) and Tra1 (**b**) by a rotation of the FRB domain. **f.** Sequence alignment of the catalytic and activation loop region of twenty-nine ΨPIKKs and PIKKs, colored by similarity as indicated. \*: residues proposed to be involved in catalysis in mTOR<sup>73</sup>. **g.** Residue Y3698 of hSAGA, equivalent to the first residue in the DFG motif of the activation loop in active kinases<sup>23</sup>, adopts a cis-peptide bond, clearly defined by the density map (contoured at 9.0 σ).



**Extended Data Fig. 10 | Buried mutations in the TRRAP module.** **a.** Overview of TRRAP (as in Fig. 5a), with the rectangle indicating the location of mutations. **b.** Magnified view of the region showing how the mutations of the two buried glycine residues (G1110W and G1158R) in two neighboring helices are not compatible with bulky sidechain residues. A translucent surface representation is shown with a cartoon model of the TRRAP HEAT repeat domain. C $\alpha$  atoms are shown as spheres. Buried residues are shown in orange and surface accessible residues are shown in red (as in Fig. 5d).

## Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

## Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	No sample size calculation was performed. Type and number of particles included in the final reconstructions were determined by resolution maximization.
Data exclusions	EM data processing: According to common standards in the field, not well aligning particles as well as particles representing different conformations were excluded. The particles were selected by resolution maximization.
Replication	EM data was collected at least three times and processing yielded similar results.
Randomization	Randomization was performed automatically by the single particle EM progressing software listed above.
Blinding	No blinding was performed as is common practice for single particle structure determination.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	SUPT7L, Santa Cruz Biotechnology cat. # sc-514548, used in WB at 1:1000. KAT2A, Cell Signaling Technology cat. #3305, used in WB at 1:1000. TADA2B, Thermo Fisher Scientific cat. #PA5-57393, used in WB at 1:2500. TBP, Abcam cat. #ab51841, used in WB at 1:2000. USP22, Santa Cruz Biotechnology cat. #sc-390585, used in WB at 1:200. TAF9B, #G2306, Homemade antibody previously created and validated in Herrera, Yamaguchi, Roelink & Tjian, ELife 2014. Used in WB at 1:1500. TAF10, Millipore Sigma cat. #MABE1079, used in WB at 1:2000. ENY2, Abcam cat. #ab183622, used in WB at 1:1000.
Validation	SUPT7L, Santa Cruz Biotechnology cat. # sc-514548: Antibody was validated by commercial supplier. From supplier: Western blotting of whole cell lysates from U-251-MG and T98G cells detect a band at ~60kDa. "SPT7L (F-5) is a mouse monoclonal antibody raised against amino acids 182-414 mapping at the C-terminus of SPT7L of human origin...SPT7L (F-5) is recommended for detection of SPT7L of mouse, rat and human origin by Western Blotting (starting dilution 1:100, dilution range 1:100-1:1000)."  KAT2A, Cell Signaling Technology cat. #3305: Antibody was validated by commercial supplier. From supplier: Western blotting of cell extracts from HeLa, NIH/3T3, C6, and COS cells show expected band at ~94kDa, and confocal immunofluorescent analysis of HeLa cells using GCN5L2 (C26A1) Rabbit mAb shows expected nuclear staining. "Monoclonal antibody is produced by immunizing animals with a synthetic peptide corresponding to the amino terminus of human GCN5L2. GCN5L2 (C26A10) Rabbit mAb detects endogenous levels of total GCN5L2 protein. The antibody does not cross-react with the related PCAF protein."  TADA2B, Thermo Fisher Scientific cat. #PA5-57393: Antibody was validated by commercial supplier. From supplier: "Western blot analysis of TADA2B in Human cell line RT-4, U-251MG sp; Human plasma (IgG/HSA depleted), Human liver tissue, and Human tonsil

tissue. Samples were probed using a TADA2B Polyclonal Antibody. Immunohistochemical staining of TADA2B in human testis tissue shows strong nuclear positivity in cells in seminiferous ducts. Immunofluorescent staining of TADA2B in human cell line A-431 shows positivity in nucleus but excluded from the nucleoli.”

TBP, Abcam cat. #ab51841: Antibody was validated by commercial supplier. From supplier: Western blot is shown for HeLa Cytoplasmic Lysate at 10ug, HeLa Nuclear Lysate at 10ug, and HeLa Nuclear Lysate at 20ug using anti-TATA binding protein TBP antibody [mAbcam 51841] - ChIP Grade (ab51841) at 5 µg/ml under reducing conditions. “This antibody gave a positive signal in the following lysates : HeLa whole cell lysate and HeLa nuclear lysate.” Validation data is also shown for Immunohistochemistry, ChIP, Flow cytometry, Immunofluorescence, and Immunoprecipitation, available at supplier webpage.

USP22, Santa Cruz Biotechnology cat. #sc-390585: Antibody was validated by commercial supplier. From supplier: Western blotting of whole cell lysates from Jurkat, HeLa, Caco-2, Ca Ski, AMJ2-C8, C6, and EOC 20 cells detect a band at ~60kDa. “USP22 (C-3) is a mouse monoclonal antibody raised against amino acids 130-176 mapping within an internal region of USP22 of human origin...USP22 (C-3) is recommended for detection of USP22 of mouse, rat and human origin by Western Blotting (starting dilution 1:100, dilution range 1:100-1:1000)...”

TAF9B, #G2306, Homemade antibody previously created and validated in Herrera, Yamaguchi, Roelink & Tjian, ELife 2014.

TAF10, Millipore Sigma cat. #MABE1079: Antibody was validated by commercial supplier. From supplier: “Clone 6TA-2B11 (a.k.a. 6TA 2B11) specifically immunostained wild-type, but not Taf10<sup>-/-</sup>, mouse embryonic blastocysts (Mohan, W.S. Jr, et al.(2003). Mol .Cell. Biol. 23(12):4307-4318)... Evaluated by Western Blotting in A431 cell lysate. Western Blotting Analysis (WB): A 1:2,000 dilution of this antibody detected TAF10/TAFII30 in 10 µg of A431 cell lysate.”

ENY2, Abcam cat. #ab183622: Antibody was validated by commercial supplier. From supplier: Western blotting of Jurkat whole cell lysate at 30ug with Anti-ENY2 antibody (ab183622) at 1/1000 dilution shows detection of band at ~12kDa.

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Cell line source(s): HeLa cells were obtained from ATCC.
Authentication	Provided by supplier. Endogenous knock-in SUPT7L cell line was validated using Sanger sequencing of the targeted locus +/- >500 bp on either side using multiple redundant primer pairs and Western blotting.
Mycoplasma contamination	Cells were routinely tested and were negative for mycoplasma contamination (microscopy fluorescence test performed by the UC Berkeley Cell Culture Facility).
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used in this study.