



HAL
open science

Structure of SAGA and mechanism of TBP deposition on gene promoters

Gabor Papai, Alexandre Frechard, Olga Kolesnikova, Corinne Crucifix,
Patrick Schultz, Adam Ben-Shem

► **To cite this version:**

Gabor Papai, Alexandre Frechard, Olga Kolesnikova, Corinne Crucifix, Patrick Schultz, et al.. Structure of SAGA and mechanism of TBP deposition on gene promoters. *Nature*, 2020, 577 (7792), pp.711-716. 10.1038/s41586-020-1944-2 . hal-03051961

HAL Id: hal-03051961

<https://hal.science/hal-03051961>

Submitted on 26 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structure of SAGA and mechanism of TBP deposition on gene promoters

Gabor Papai^{1,2,3,4}, Alexandre Frechard^{1,2,3,4}, Olga Kolesnikova^{1,2,3,4}, Corinne Crucifix^{1,2,3,4}, Patrick Schultz^{1,2,3,4}# and Adam Ben-Shem^{1,2,3,4}#

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Integrated Structural Biology, Illkirch, France, Equipe labellisée Ligue Contre le Cancer.

²Centre National de la Recherche Scientifique, UMR7104, Illkirch, France.

³Institut National de la Santé et de la Recherche Médicale, U1258, Illkirch, France.

⁴Université de Strasbourg, Illkirch, France.

Corresponding authors. Patrick Schultz (patrick.schultz@igbmc.fr) and Adam Ben-Shem (adam@igbmc.fr)

Abstract

SAGA is a 19-subunit complex that stimulates transcription via two chromatin-modifying enzymatic modules and by delivering TBP to nucleate the pre-initiation complex on DNA, a pivotal event in protein gene expression. Here we present the structure of yeast SAGA with bound TBP where the core is resolved at 3.5 Å resolution (0.143 FSC). It elucidates the intricate network of interactions that coordinate the different functional domains of SAGA and resolves an octamer of histone-fold domains at the core of SAGA. This deformed octamer, deviates considerably from the symmetrical analogue in the nucleosome and is precisely tuned to establish a peripheral site for TBP where steric hindrance represses binding of spurious DNA. With complementary biochemical analysis our structure points to a mechanism for TBP delivery and release from SAGA that requires TFIIA and whose efficiency correlates with the affinity of DNA to TBP. We provide the foundations for understanding specific TBP delivery onto gene promoters and the multiple roles played by SAGA in regulating gene expression.

Main Text

Transcription of protein-coding genes begins with the formation of a pre-initiation complex (PIC) composed of RNA polymerase II and several general transcription factors (TFs)^{1,2}. PIC assembly is nucleated by loading the TATA-binding protein (TBP) onto promoter DNA^{3,4}, a focal point for regulated gene expression⁵. Two multi-protein co-activator complexes, TFIID and SAGA can deliver TBP to the gene promoter^{6,7} and are required for global gene expression in yeast^{8,9}.

The 1.6 MDa SAGA complex also stimulates transcription via its two chromatin-modifying enzymatic activities. It is composed of 19 subunits organized in four modules with distinct functions¹⁰: a histone acetyl transferase (HAT) module¹¹, a histone deubiquitinase (DUB) module¹², the 430 kDa Tra1 subunit which serves as a docking platform for transcriptional factors that recruit SAGA to activating DNA sequences upstream of the promoter¹³, and a 10-subunit central module that is physically connected to all other modules and is also responsible for recruiting TBP to SAGA^{10,14} (Extended Data Fig. 1).

Structures of several individual isolated components of SAGA, notably the DUB module¹⁵⁻¹⁷, the catalytic subunit of the HAT module^{18,19} and Tra1^{20,21} were obtained. Previous electron microscopy (EM) studies positioned Tra1 in one lobe of the SAGA complex but all other subunits appeared together as one main lobe that could only be described at low resolution²²⁻²⁴.

The central module of yeast SAGA (Extended Data Fig. 1) shares five TBP associated factors (TAFs) with TFIID. Both complexes also harbor multiple histone-fold (HF) domains that associate into heterodimers and higher-order oligomers²⁵⁻²⁷. The central module is therefore expected to be structurally related to TFIID suggesting TBP binding in both complexes has some common features. Progress in defining the architecture of TFIID has been recently reported however the machinery responsible for TBP binding and delivery remained poorly resolved^{28,29}.

We report the 3.5 Å resolution structure of the full SAGA complex bound to TBP, determined from cryo-EM maps that show clear density for the majority of protein side chains. Our structure describes in atomic details the network of interactions that coordinate the functional domains of SAGA as well as a machinery precisely geared to binding TBP and preventing aberrant transcription initiation. With complimentary experimental findings our structure leads to a mechanism of TBP delivery and release from SAGA that depends on TFIIA and the affinity of promoter DNA to TBP, providing new insights to PIC assembly regulation.

Overall architecture of SAGA

We reconstituted the complex between purified SAGA from the yeast *Komagataella phaffii* (*Pichia pastoris*), full length TBP and ubiquitin aldehyde and collected a dataset of over 1 million particles. Image analysis resulted in a map of the full SAGA complex at 3.8 Å resolution calculated from 338,850 selected particles (Fig. 1a; Extended Data Fig. 2, Extended Data Fig. 3 and Table 1). Focused refinement of the main lobe and of Tra1 further improved their average resolution to 3.5 Å, with many regions reaching as high as 3 Å. Hence, Tra1 and the central module (subunits Spt7, Spt3, Ada1, Spt20 and the five TAFs) as well as the domains of Sgf73 that anchor the DUB to the central module show clear densities for the majority of side chains (Fig. 1b and Extended Data Fig. 4).

At the periphery of SAGA, exposed next to Spt3, we identified the distinctive saddle shaped density of the conserved core of TBP (cTBP), absent in a separate 5 Å resolution map of SAGA lacking TBP (Extended Data Fig. 5, Extended Data Fig. 6a). Local classification revealed sufficient secondary structural features to guide the precise positioning of cTBP (Extended Data Fig. 6b). Site-specific cross-linking, as well as a large body of well documented mutations, resolved any ambiguity in orientation arising from the pseudo-symmetric fold of cTBP³⁰⁻³². A disc shaped density neighboring TBP was assigned to the WD40 protein Spt8 which was shown to interact with TBP³³ (Extended Data Fig. 6c,d). At the opposite pole of the main lobe, the DUB crystal structure fits into and occupies a large but not well resolved density that is nearly completely detached from and independent of SAGA. The DUB module is connected to the main lobe by only a narrow, elongated density that contacts SAGA's surface at a single site, proximal to the Spt20 SEP and the TAF5 LisH domains (Fig. 1b and Extended Data Fig. 1). The second enzymatic module of SAGA which harbors the HAT activity adopts multiple conformations and orientations. Accordingly, the bulk of this module is visible only as an elongated featureless density at the tip of our map. Nevertheless, we could identify its docking site on SAGA where two helical domains, attributable to Ada3 according to 2D predictions, cross-linking mass spectrometry (XL/MS) and analysis of internal domain deletions¹⁴, lie at the surface of the Taf6 HEAT repeats domain.

Taken together, the structure described here defines the complete architecture of SAGA and resolves at near atomic resolution almost all protein side chains comprising SAGA's core and Tra1. It elucidates how the central module, which acts as a scaffold, coordinates an activator-binding platform, two peripheral enzymatic modules and the TBP binding machinery.

Taf5 - an orchestrating hub of the main lobe.

The main lobe is dominated by the 7-blade WD40 propeller of Taf5 (Fig. 2). This domain occupies a central position, serving as a structural hub connecting all major SAGA functions. The flat top face of the propeller disc clamps together the Taf5 NTD and the TAF6 HEAT repeats which bind the HAT module. The opposite WD40 bottom face holds the histone-fold machinery that ultimately binds TBP and also connects to the DUB module and to Tra1.

The positions of the Taf5-NTD and Taf6 HEAT repeats are dramatically different in SAGA compared to TFIID²⁸, as these domains assume new roles in establishing connections to the enzymatic DUB (LisH extension of Taf5-NTD) and HAT (Taf6) modules (Fig. 2a, Extended Data Fig. 7). Compared to TFIID, the Taf5-NTD in SAGA makes almost a 180 degrees turn around the WD40 domain and the Taf6 HEAT repeats, remote in TFIID, are relocated to the vicinity of the WD40 top face in SAGA. It is also noteworthy that in contrast to SAGA which contains a single copy of Taf5 and Taf6, TFIID harbors two copies of these subunits that are held together by homodimerisation of Taf5-NTDs and Taf6 HEAT repeats. As a result of all these variations and of additional interactions with SAGA specific subunits, the Taf5 WD40 environment is very different in SAGA as compared to TFIID.

The ability of Taf6 HEAT repeats to occupy three different positions in SAGA and TFIID relative to its histone fold is afforded by a 136 residue-long linker between these two domains. Our structure provides a nearly complete description of this linker that partially encircles the Taf5 WD40 domain (Fig. 2b). The Taf6 linker position is stabilized through several interactions along its trajectory. Most notably Taf6 linker residues T135-L141 add a beta-strand to the last blade of the WD40 propeller forming an inter-protein beta sheet.

WD40 domains are recognized as potential hubs that mediate protein-protein interactions. The Taf5 WD40 in SAGA is a staggering manifestation of this capacity as it associates with at least 11 different protein domains (Fig. 2c). Hence Taf5 is a centrally located master hub that orchestrates the architecture of the entire main lobe.

A deformed histone fold octamer defines the TBP binding site

An octamer consisting of four HF pairs arranged in a disc-shaped spiral, akin to the nucleosomal histone octamer, was proposed to reside in TFIID and SAGA³⁴⁻³⁶. Our structure offers the first atomic model of the full HF octamer that resides at the core of SAGA and most probably with minor variations in the TBP-binding lobe of TFIID (Fig. 3a). It reveals a striking disparity between HFs interactions in SAGA as compared to the nucleosome, particularly regarding the incorporation into SAGA of the TBP binding subunit Spt3, a homologue of the TFIID Taf11-Taf13 HF pair³⁵.

In the nucleosome the interaction between neighboring histone pairs is mediated by a tightly packed four-helix bundle composed of helices $\alpha 2$ and $\alpha 3$ of one histone from each pair³⁷. The SAGA HF pairs are organized along the spiral in the following order: Taf6-Taf9, Ada1-Taf12, Spt7-Taf10 and Spt3 which harbors two histone folds, one at each of its ends, nSpt3-HF and cSpt3-HF. While all other HF pairs are oriented as in the nucleosome (Fig. 3b), the Spt3 pair is tilted by 20 degrees compared to its analogous nucleosome H2A-H2B histone pair (Fig. 3c). This tilt prevents participation of the Taf10-

α 3 helix in the interaction, shears the four-helix bundle and therefore nearly completely frees Spt3 from its association with the HF octamer. The remaining contact is facilitated by a slight tilt in Taf10 and maintained only by stacking between the highly conserved nSpt3-F72 and Taf10-R120 residues as well as hydrogen bonds between nSpt3-R60 and Taf10 residues (carbonyl atom of I121 and side chain of S125) (Fig. 3d). To secure incorporation of Spt3 into SAGA the C-terminal tail of Spt3 inserts into the center of the spiral disc, a void volume in the nucleosome, makes a near complete circle and forms single bonds with each of the other histone folds in SAGA. It is noteworthy that the 40 residue Spt3 tail is stabilized all along its path as indicated by the clear density until the very C-terminus (Fig. 3e).

The Spt3 deviation from the canonical octamer does not originate from a single structural determinant. The atypical sequences of the histone folds (in particular Taf10 and Spt3), the long Taf10 loop between its α 2 and α 3 helices, the inherent asymmetry arising from the 8 different histone-fold proteins involved, all contribute. In this respect, it is noteworthy that the deviation from the nucleosome pattern is already apparent at the preceding four-helix bundle, linking Taf12-Ada1 to Spt7-Taf10, where the association lacks almost entirely hydrophobic packing. We propose that this assembly of histone folds, deviating considerably from the symmetric arrangement in the nucleosome, is geared specifically towards a delicate balance between a rigid Spt3 that is firmly embedded into the octamer and an overly flexible Spt3 that associates with SAGA through loops alone. Below we examine these ideas within the context of TBP binding and release from SAGA.

TBP binding and release machinery

SAGA presents TBP at the extremity of the main lobe embracing it by binding to two opposite ends of its surface and orienting the DNA-binding cleft of TBP towards the main body of SAGA (Fig. 4a). The principle component in coupling TBP to SAGA and determining its position is Spt3. The C-terminal stirrup of TBP is the primary binding site for Spt3, similar to the interaction between TBP and TFIIB³⁸. In SAGA, the stirrup is buried nearly completely and forms multiple contacts within a large pocket created by helices α N and α 2 from cSpt3-HF and by helix α C that is joined to nSpt3-HF via a conserved two amino acid rigid turn (Fig. 3d, Fig. 4b). Additional contacts with TBP are probably established by a stretch of residues just N-terminal to helix α N of cSpt3-HF as well as by the preceding helix (α Loop), part of the 130 amino acids loop that connects the two histone folds in Spt3, that lies in parallel to helix H1' of TBP. The association of the C-terminal half of the pseudo-symmetric cTBP with Spt3 is in line with site-specific short-range photo-cross-linking between Spt3 and TBP-S184³². Indeed, accommodating cTBP N-terminal half by Spt3 is unfavorable as it would have the negatively charged D81 facing a negatively charged patch on Spt3 (E203-E202) and would place the bulky charged R98 where the C-terminal stirrup L189 is tightly packed against a hydrophobic crevice (Extended Data Fig. 8).

In contrast to Spt3, we find that Spt8 is not part of the central module (Fig. 4a). It is flexibly tethered via a fuzzy density to the outward tips of two long nearly parallel helices, extensions of each end of the Spt7 histone fold, that reach 50 Å away from the main body of SAGA. The fuzzy density corresponds, at least in part, to a 30 residue long unstructured region in the C-terminus of Spt7, whose deletion results in loss of Spt8³⁹. It is important to note that mass spectrometry and SDS-PAGE analysis²¹ exclude the possibility that either sub-stoichiometric amounts of Spt8 or cleavage of Spt7 account for the lower-resolution in this region of the map. The connection of Spt8 to the N-terminal half of cTBP is very flexible and appears to play a secondary role in positioning TBP. This finding is consistent with deletion mutants of Spt3 in yeast, showing sharp decline in global levels of newly synthesized mRNA, whereas impairment of Spt8 showed practically no effect⁸. We suggest Spt8 plays an auxiliary role in tuning the orientation of TBP and in competing with other TBP-associated proteins,

such as TFIIA or Mot1, whose binding surfaces on TBP overlap partially with that of Spt8⁴⁰⁻⁴². The flexible tethering of Spt8 to SAGA core contributes to small movements of TBP that are reflected in a lower local resolution. This “breathing” of TBP might be important to transiently allow initial access to DNA or regulatory factors.

The space between TBP and the main lobe of SAGA is just wide enough to accommodate a double-stranded DNA bent by TBP (Fig. 5a). However, the path of DNA is obstructed by structural elements located distally to the TBP DNA binding cleft, notably the peripheral ring of the Tra1 HEAT repeats, as well as the nSpt3-HF helix α C, a major component of the pocket that envelopes the C-terminal TBP stirrup. Gel shift (Fig. 5b) and pull-down assays (Fig. 5c and d, Extended Data Fig. 9) of DNA binding to SAGA-TBP demonstrate that this mild steric hindrance is none-the-less sufficient to impede binding of even high-affinity DNA containing a consensus TATA box (Fig. 5b lane 10, Fig. 5c and d).

The general transcription factor TFIIA enhances DNA binding to TBP, competes with Spt8 for occupying the N-terminal half of cTBP and is found in pre-initiation complexes nucleated by SAGA⁴³⁻⁴⁵. When a consensus TATA-box containing DNA was incubated with the SAGA-TBP complex together with TFIIA, TBP was released from SAGA as a ternary DNA-TBP-TFIIA complex (Fig. 5b lane 3). The binding of DNA to TBP is linked with the release of TBP as we did not observe any DNA associated with SAGA-bound TBP. Indeed, the release of TBP from SAGA can now be understood as essential for the subsequent steps in assembling the transcription pre-initiation complex because the clasping of the TBP C-terminal stirrup by Spt3 and TFIIB is mutually-exclusive. TBP was also detached from SAGA, though less efficiently, when a DNA bearing the TATA element containing two mismatches was used (Fig. 5b lanes 5-6), whereas no released TBP was observed when the DNA did not contain any TATA or TATA-like elements (Fig. 5b lanes 7-8). These experiments show that the detachment of TBP from SAGA correlates with the affinity of TBP for its DNA substrate.

Our structure suggests that the TFIIA-stimulated mechanism for the concerted DNA binding and TBP release from SAGA includes the following steps: TFIIA first displaces Spt8 from TBP; it then promotes an initial binding of DNA to TBP thus shifting Spt3 helix α C that obstructs the path of DNA. Importantly, as helix α C is firmly joined to helix α 3 in nSpt3-HF, this crucial step is made possible by the fact that the deformed octamer does not immobilize this HF via a tight four helix bundle. Shifting helix α C breaks the pocket that clasps the C-terminal stirrup; TBP can then tilt, letting DNA pass in the gap between Tra1 and SAGA main body, completing the binding to DNA. Final release of TBP from SAGA might still require a further Spt3 conformational change. Spt3 alone is unable to bind TBP³³ and it is tempting to speculate that by shifting helices α C- α 3 of nSpt3-HF, partially loosening Spt3 ties with SAGA, a conformation akin to isolated Spt3 ensues, which favors TBP detachment. Following the release of TBP and DNA, helix α C can regain its position allowing Spt3 to restore its TBP-binding conformation.

It has been suggested that in TFIID the association of TBP with spurious DNA, which could lead to aberrant gene activation, is prevented through a helical protein domain that mimics TATA-DNA and binds at the DNA cleft of TBP^{46,47}. Remarkably our maps do not show any significant density attached at the DNA binding cleft of TBP. Instead, our structure reveals a different mechanism for warding off non-specific DNA based on limited steric hindrance, distal to the DNA cleft in TBP. Only the synergy of a cognate TATA or TATA-like DNA with TFIIA can overcome this hindrance.

Firmly tethered but flexible enzymatic modules

The DUB and HAT modules are highly dynamic (Fig. 1; Extended Data Fig. 10), permitting the exploration of a large conformational space in search of their chromatin substrates and the structure shows how these modules maintain however a robust binding to SAGA. SAGA DUB module consists of the catalytic subunit Ubp8 as well as Sgf11, Sus1, and the Sgf73 N-terminal end. The C-terminal part of Sgf73 was implicated in attaching DUB to SAGA and deletion-analysis narrowed down the anchoring region to residues 350-400 in *S. cerevisiae* (227-277 in *K. phaffii*) since removal of this region leads to dissociation of an intact DUB module from SAGA^{14,48}. We find that these residues form an elongated domain composed of two helices, Sgf73-h1 and Sgf73-h2, connected by a stretched loop and embedded within the central module (Fig. 6a). This DUB anchoring domain traverses the core, entering proximal to Taf9-HF and exiting next to the Spt20-SEP domain (Extended Data Fig. 1). Sgf73-h2 is well resolved and all side chains could be clearly discerned, indicating its stable interaction with a chamber-like structure composed of several helices from the Spt20-SEP domain and from an elongated domain of Ada1, located N-terminal to the histone fold. The chamber consists exclusively of SAGA-specific elements with no contribution from subunits shared with TFIID. Hence, at the core of SAGA a dedicated environment was created, partially occupied in TFIID by the Taf5-NTD, to house the helix that anchors the DUB module into SAGA. The DUB module is tethered to SAGA's core through the 164aa linker connecting the anchoring domain to the N-terminal end of Sgf73 which is part of the DUB. This linker includes a SCA7 domain but counts in addition roughly 90 amino acids predicted unstructured. We find little trace of this linker in the map except for a short stretch contacting the Spt20-SEP and Taf5-LisH domains. However, the linker was shown to cross-link with residues located at the surface of SAGA allowing us to delineate its approximate path (Extended Data Fig. 11).

In contrast to the DUB, no parts of the HAT module are embedded within SAGA core (Fig. 1b). The HAT densities emerge, harboring two distinct helical domains attributable to subunit Ada3 (Extended Data Fig. 12), at the surface of the Taf6 HEAT repeats that is used for homo-dimerization in TFIID (Fig. 6b).

Thus, the two enzymatic modules of SAGA adopt very different strategies to remain bound to SAGA while maintaining their freedom of movement. The DUB is a stably folded structure¹⁷ that gains independence through an unstructured long linker which forms little contact with SAGA surface until it connects to an extension firmly embedded in SAGA core. The HAT, on the other hand, is a flexible structure by-itself¹⁹, with several sub-modules as well as a large proportion of intrinsically unfolded sequences, and it binds only to the surface of SAGA core employing two helical domains of Ada3.

An intricate network of unstructured protein elements interconnects the functional domains

Three distinct bridges serve to physically and possibly functionally couple Tra1 with the central module. All bridges are unidirectional, reaching out from the central module into Tra1 density, and all bind to Tra1 at, or proximal to, its FAT domain.

The first bridge is established by a highly dispersed and extended domain of Spt20 that originates as a strand joining the Spt20-SEP beta-sheet followed by a helix-turn-helix that also interacts extensively with the Spt20-SEP domain (Fig. 6c). It continues as a ~70 residue loop, touching Taf5-WD40, Taf9, Taf12, Ada1 and Spt7 as it traverses the surface of SAGA core on its 160 Å-long way to Tra1. There it forms a helix that associates with the surface of the FAT domain and is followed by a second helix which is clamped within a concave structure established by residues 2700-2753 in Tra1.

This second helix of Spt20 on Tra1 corresponds to a region (residues 325-340) that is denoted as HIT domain and whose deletion prevents assembly of Tra1 into SAGA⁴⁹.

An exceptionally long loop that precedes the histone fold of Taf12 is responsible for establishing the second bridge (Fig. 6d). Residues 430-474 from this loop trace a lasso shaped thread at the surface of the Tra1 FAT domain within the deep groove formed between the first and second TRD. The third bridge is formed by the loop connecting the second and third helices of cSpt3-HF as it infiltrates between HEAT repeats 46 and 47 of the Tra1-ring (Fig. 6e, Extended Data Fig. 13).

Several lines of evidence indicate allosteric regulation pathways between Tra1 and the enzymatic modules⁵⁰⁻⁵². A point in case is the finding that during the cellular assembly of SAGA in *S. pombe*, Tra1 controls the incorporation of DUB although the two modules do not interact directly⁴⁹. Our structure shows that it is conceivable that Tra1 induces, through the Spt20 bridge-forming domain, a conformational change in Spt20-SEP that would open the chamber to let the DUB anchoring domain enter. Furthermore, several observations support the notion that activator binding to Tra1 affects the functional ends at the upper lobe. The fact that Spt3 is directly connected to Tra1 provides a possible means for activators to regulate TBP binding to SAGA.

We observe several protein stretches which lack secondary structure elements but contact multiple subunits and contribute considerably to an intricate intertwined network of interactions within SAGA main lobe. For example, the C-terminal tail in Taf9 and the short loop preceding it, contact Spt3 C-terminal tail, Spt20 bridge domain, Taf12-HF, Ada1, Sgf73 anchoring domain, Taf6-HF, Spt7-HF, Taf5-WD40, Taf5-NTD and Taf10-HF (Extended Data Fig. 14). This network of interactions is necessary for stabilizing an octamer which could not be reconstituted *in vitro* from its constituent histone-folds and underlies the effects of some mutations on functions not directly involving the altered proteins^{24,50,53}.

Conclusion

The high-resolution description of SAGA unravels how the central module acts as a scaffold that coordinates an activator interaction platform, two enzymatic modules as well as TBP binding and release. Our structure reveals that a deformed histone-fold octamer, aided by components of Spt7 and Spt8, is finely tuned to define a TBP docking site where distal steric hindrance represses DNA binding thus preventing unwanted transcription initiation. However, with the assistance of TFIIA, TBP can productively bind DNA to form a ternary TBP-TFIIA-DNA complex that is released from SAGA. The efficiency/probability of TBP detachment is commensurate with the affinity of TBP to the DNA sequence.

These findings imply a short residency time for SAGA on the minimal promoter in accordance with genome wide analyses that locate SAGA mainly at upstream activating sequences (UASs)^{8,54}. The peripheral position of TBP and its binding mode in SAGA also suggests that transcription regulatory factors such as NC2 or Mot1 could gain access to TBP with relative ease. These results are consistent with a highly dynamic scenario for the control of TBP turnover entailing regulatory factors, the affinity of TBP to promoter DNA as well as chromatin markers and sequence specific activators that recruit SAGA to the UAS⁵⁵.

Sequence homology strongly suggest that the histone-fold machinery in TFIID assumes a similar function which implies that the mechanics of depositing TBP on DNA by the two complexes share common features. However several crucial aspects of TBP binding and delivery vary between the two complexes including the capacity of TFIID subunits Taf2 and Taf4 to bind promoter DNA, the ability of Taf1-NTD to interact with the DNA binding interface of TBP and the stronger affinity of Taf11-13, compared to Spt3, for TBP⁵⁶. Hence access of transcription regulatory factors to TBP as well as

residency time at the minimal promoter are expected to be significantly different between TFIID and SAGA.

Two complementary TBP delivery systems have been developed and conserved during eukaryotic evolution. The present structure will guide future work aiming at elucidating how the division of labor between SAGA and TFIID in nucleating the PIC is determined for specific genes. Further studies are also required to decipher the combinatorial impact on SAGA recruitment exerted by sequence-specific transactivators, histone modifications, nucleosome deplete regions and chromatin remodelers upstream of the promoter as well as to understand the complementary role played by the enzymatic activities of SAGA, which are absent in TFIID, in PIC formation.

Methods

Purification of SAGA.

The SAGA complex was purified from nuclear extracts of the budding yeast *Komagataella phaffii* (known also as *Pichia pastoris*) using the 38 amino acid streptavidin-binding peptide (SBP) affinity tag that was fused to the C-terminus of the endogenous Sgf73 subunit (Extended Data Fig. 1). We introduced several modifications into our original procedure (Sharov et al. 2017) with the combined effect of increasing the yield by 3 times. These improvements enabled us to reduce the harmful effects of proteases and to keep SAGA more concentrated along the different steps of the procedure.

Two liters of yeast cells (4 flat-bottom flasks with 0.5L in each) were grown at 24°C with glycerol as carbon source and harvested when OD_{600 nm} reached 16-18. Cells were washed in ice-cold water and then treated with 10 mM DTT. The cell wall was digested by addition of lyticase and spheroplasts were pelleted at 5,500 g for 20 min. All further steps were performed at 0 to 4°C. Protease inhibitors were added to all buffers. Spheroplasts were washed twice in 1.1 M Sorbitol and were then disrupted by suspension in a hypotonic buffer (15-18% Ficoll 400, 0.6 mM MgCl₂, 20 mM K-phosphate buffer pH 6.6) using a ULTRA-TURRAX disperser. Sucrose (0.1 M) and MgCl₂ (5 mM) were then added. Nuclei (and some debris) were pelleted at 33,000 g for 37 min, resuspended in a wash buffer (0.6 M Sucrose, 8% PVP, 1 mM MgCl₂, 20 mM phosphate buffer pH 6.6) and pelleted again at 34,000 g for 50 min. Nuclei were resuspended in low ionic extraction buffer (40 mM HEPES pH 8.0, 20% sucrose, 8 mM MgCl₂, 5-6 mM DTT) with 30 strokes using a tight pestle in a dounce homogenizer. Very few molecules of SAGA are liberated under the low ionic conditions leaving time for protease inhibitors to bind their targets before SAGA is extracted.

After 20 minutes of incubation, 300 mM NaCl, 2 mM CaCl₂ and 150 ul of alpha-Amylase solution (MegaZyme) were added. Following another 30 minutes of incubation, debris were precipitated at 33,000 g for 38 min. The supernatant was collected and 1-2 % PEG 20,000 as well as additional 5 mM MgCl₂ were added in order to precipitate some remaining organelles and membrane parts by a short centrifugation step at 33,000 g for 10 min. The increase in magnesium concentration (final 13 mM) proved crucial to prevent precipitation of SAGA at this step. The PEG 20,000 concentration was then increased to 5.8% and SAGA precipitated in a second short centrifugation step. The pellet was solubilized in a minimal volume and avidin was added to block endogenously biotinylated proteins. The suspension was incubated with streptavidin beads for 4 h in buffer A (40 mM HEPES pH 8.0, 250 mM Sodium Chloride, 10% sucrose, 2 mM MgCl₂, 2 mM DTT) washed 5 times and eluted with buffer A containing 10 mM biotin. The eluate was concentrated with Millipore Amicon-Ultra (50 kDa cut-off) and spun in a 10-30% sucrose gradient with buffer B (20 mM HEPES pH 8.0, 150 mM Potassium acetate,

2 mM DTT, 6 mM MgCl₂) in rotor SW60 (38,300 rpm for 13.5 h). SAGA was fractionated at approx. 25% sucrose and concentrated with Amicon-Ultra to ~ 2 mg/ml.

Protein purification.

Full-length yeast *S. cerevisiae* TBP was expressed as N-terminal 6-HIS-GFP fusion protein in *Escherichia coli* strain BL21(DE3)RIL. Cells were grown in LB medium with ampicillin until an OD_{600 nm} of 0.7 was reached. Expression was induced with 1 mM IPTG for 3 h at 37°C. The cells were collected by centrifugation and resuspended in lysis buffer (20mM Hepes-KOH (pH7.5), 0.5M NaCl, 10% glycerol). Cells were treated with 1 mg/ml lysozyme (Sigma) for 30 min on ice, and then lysed by sonication after the addition of cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche) and PMSF to a final concentration of 1mM. The resulting lysate was clarified by centrifugation and incubated with equilibrated cOmplete™ His-Tag Purification Resin (Roche) for 1 h at 4°C. After the beads were collected, they were washed three times with lysis buffer supplemented with 5mM imidazole, and the bound protein was eluted with 20mM Hepes-KOH (pH7.5), 0.2M NaCl, 10% glycerol supplemented with 250 mM imidazole. The recombinant protein was further purified on a Heparine HiTrap 1ml column (GE Healthcare) equilibrated with 20mM Hepes-KOH (pH7.5), 0.2M NaCl, 20% glycerol and eluted with a gradient from 200 mM to 1M NaCl. Fractions were analyzed by Coomassie blue staining, pooled, and concentrated in Amicon Ultra with 10,000 (10K)-molecular-weight cutoff (Millipore) to a final concentration of 3 mg/ml. DTT (5mM) was added to all buffers immediately before use.

The yeast *S. cerevisiae* TFIIA was expressed in *E. coli*, purified from inclusion bodies and reconstituted as described earlier⁵⁷.

Gel shift DNA binding assays.

DNA probes were obtained by annealing of equimolar amounts of complementary oligonucleotides (Eurofins Genomics for short DNA, SIGMA for long DNA) at a final concentration of 25 μM in 20mM Hepes-KOH, 50mM NaCl by heating the mixture to 100°C for 5 min and cooling slowly down to room temperature. To estimate the quality of annealing, DNA fragments obtained were loaded on 12% PAGE and analyzed using a Typhoon FLA9500 Imager (GE Healthcare) to follow Cy5 fluorescence signal. No Cy5-labeled single stranded DNA was detected (ED Fig 8e). To further ascertain that we remove all effects of single stranded DNA, all DNA samples were treated with Exonuclease I (NEB). Binding assays (10μl) contained 3 mM Tris-HCl (pH 8.0), 6mM Hepes-KOH (pH8.0), 45 mM KAc, 42 mM KCl, 5 mM MgCl₂, 3% glycerol, 6% sucrose, DNA probe (Cy5-labeled), 0.01% NP40, 25 μg/ml bovine serum albumin (BSA) (NEB), 1 mM DTT, and proteins as indicated. TBP was incubated first with four-fold excess of SAGA complex for 1h at 4°C. Under these conditions almost all TBP was bound to SAGA. Then DNA and TFIIA were added, and the reaction mixtures were incubated for 20 min at room temperature before they were loaded onto agarose-polyacrylamide native gels (1% agarose, 1,3% acrylamide with 1x Tris-glycine buffer, 5mM MgCl₂ in the gel and running buffer). Final reaction mix contained 0.075 μM GFP-TBP, 0.3 μM SAGA, 0.075 μM DNA and 0.15 μM TFIIA complex. Binding assays with long (100 bp) DNA fragments were done under similar conditions except that a 3 fold excess (with respect to SAGA) of long non-labeled TATA-less_L DNA was also added because SAGA was shown to bind long-DNA non-specifically⁵².

Gels were analyzed using a Typhoon FLA9500 Imager to follow Cy5 fluorescence signal. TBP-DNA and TBP-TFIIA-DNA migrate similarly in native agarose-polyacrylamide gels due to the limited resolving power of these gels for relatively low molecular-weight complexes. We ascertained the composition

of the shifted bands by mass-spectrometry (ED Fig. 8d). Sequences of DNA molecules used: “TATA” = yeast *CYC1* TATA 5'-Cy5-TGCTCTGTATGTATATAAACTCTTG-3', “TATA-like” = *CYC1* TATA with two mutations decreasing the affinity to TBP 5'-Cy5-TGCTCTGTATGTAGAGAAACTCTTG-3', “TATA-less” = non-specific DNA, *CYC1* UAS 5'- Cy5-GGCCGGGGTTTACGGACGATGGCAG-3', “TATA_L” = *Pichia pastoris* pGAP promoter fragment TGTCATGAGATTATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCTT TCCCAATTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTA, “TATA-like_L” = TATA_L with two mutations decreasing the affinity to TBP TGTCATGAGATTATTGGAAACCACCAGAATCGAAGAGAAA AGGCGAACACCTTCCCAATTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTA, “TATA-less_L” = non-specific DNA GACTCATGTCATGAGATCATTGGACACCACCAGAATCGCGTATCGAAGGCGAACACCTGTCTCA CGTCTGGTGTCTCCTGACGCACAGACTTCGAACGTA.

Pull-down DNA binding assays.

Binding assays (20µl) contained 3 mM Tris-HCl (pH 8.0), 5mM Hepes-KOH (pH8.0), 35 mM KAc, 46 mM KCl, 5 mM MgCl₂, 3% glycerol, 6% sucrose, 0.01% NP40, 25 µg/ml bovine serum albumin (BSA) (NEB), 1 mM DTT, proteins and DNA probe (Cy5-labeled “TATA” DNA) as indicated. TBP was incubated first with four-fold excess of SAGA complex for 1h at 4°C. Then DNA and TFIIA were added (as indicated), and the reaction mixtures were incubated for 20 min at room temperature. Final reaction mix contained 0.075 µM GFP-TBP, 0.3 µM SAGA (with Sgf11 subunit fused to mCherry), 0.075 µM Cy5-DNA and 0.15 µM TFIIA complex. Ten µl of streptavidin-sepharose beads (equilibrated with the binding buffer) were added to each reaction mix and tubes were incubated for 2 hours at 4°C with shaking using ThermoMixerC (Eppendorf, shaking frequency 1220 rpm). Each reaction mix containing beads was transferred into Ultrafree-MC centrifugal filter unit and centrifuged 5 min at 12000 g at 4°C. Beads with retained protein or DNA-protein complexes, and the flowthrough with unbound material were analyzed using PHERAstar Plus microplate reader (BMG LABTECH) to quantify the fluorescence signal of GFP (corresponding to TBP), mCherry (corresponding to SAGA complex) and Cy5 (corresponding to DNA). A complementary experiment with reverse order of incubations was also carried out. Hence, TBP was first incubated with Cy5-labeled DNA and TFIIA for 20 min at room temperature, then four-fold excess of SAGA complex was added and reactions incubated for 1h at 4°C. Experiments were repeated three or four times and *error bars* represent *standard deviation* of the mean

Cryo-EM sample preparation and data acquisition.

Concentrated SAGA was incubated over-night on ice with an excess of Ubiquitin-aldehyde (UbAl) at a molar ratio of 1:4. Complete binding and inhibition was ascertained by following the activity of SAGA using Ubiquitin-AMC (Enzo). SAGA-UbAl complexes were precipitated with PEG to remove sucrose (Sharov et al. 2017), suspended at a concentration of 1 mg/ml and incubated over-night on ice with a 5:1 molar excess of full length *S. cerevisiae* TBP. SAGA-TBP was then diluted to 0.2 mg/ml and cross-linked with 0.1% glutaraldehyde for 30 min on ice. About 3 µl of sample was applied onto a holey carbon grid (Quantifoil R2/2 300 mesh) rendered hydrophilic by a 90 s treatment in a Fischione 1070 plasma cleaner operating at 30% power with a gaz mixture of 80% argon: 20% oxygen. The grid was blotted for 1.5 sec (blot force 8) and flash-frozen in liquid ethane using Vitrobot Mark IV (FEI) at 4°C and 95-100% humidity. To verify that the occupancy of GFP-TBP on SAGA was approaching 100%, we loaded the cross-linked complex on a superose 6 10/300 GL column (GE Healthcare) and compared the fluorescent signal at the peak fraction with that of cross-linked GFP-TBP of known concentrations. The concentrations of SAGA were estimated by UV absorbance.

Images were acquired on a Cs-corrected Titan Krios (Thermo Fisher) microscope operating at 300 kV in nanoprobe mode using the serialEM software for automated data collection⁵⁸. Movie frames were recorded on a 4k × 4k Gatan K2 summit direct electron detector after a Quantum Ls 967 energy filter using a 20 eV slit width in zero-loss mode. Images were acquired in super-resolution mode at a nominal magnification of 105,000, which yielded a pixel size of 0.55 Å. Forty movie frames were recorded at a dose of 1.32 electrons per Å² per frame corresponding to a total dose of 52.8 e/Å², but only the last 38 frames were kept for further processing.

Image processing.

Movie frames were aligned, dose-weighted, binned by 2 and averaged using Motioncor2⁵⁹ to correct for beam-induced specimen motion and to account for radiation damage by applying an exposure-dependent filter. Non-weighted movie sums were used for Contrast Transfer Function (CTF) estimation with the Gctf⁶⁰ program, while dose-weighted sums were used for all subsequent steps of image processing. After manual screening, images with poor CTF, particle aggregation or ice contamination were discarded. Particles were picked using crYOLO (doi: <https://doi.org/10.1101/356584>). These datasets were analyzed in Relion 3.0⁶¹, cryoSPARC⁶² and cisTEM⁶³ according to standard protocols. Briefly, three rounds of reference free 2D classification of the individual particle images were performed in Relion to remove images, like ice contaminations or deformed particles, which failed to enter into high-resolution class averages. The selected images were aligned against a low-pass filtered starting model²¹. Three rounds of 3D classification with increasing regularization parameter T were performed and classes showing high resolution features were selected for the subsequent steps. These images were imported, and refined in cryoSPARC v2. Three-dimensional classification of the entire dataset could not clearly separate distinct conformations of SAGA complex. Therefore, we carried out a focused refinement of the separate lobes using the masked lobes as references. Tra1 lobe was refined with a mask covering the Tra1 densities using the “Local refinement” module. The structure of the main lobe was obtained by using signal subtraction, where the densities corresponding to the Tra1 lobe were subtracted from the images. This operation was followed by local refinement with a mask covering the main lobe densities. Global resolution estimates were determined using the FSC = 0.143 criterion after a gold-standard refinement. Local resolutions were estimated with ResMap⁶⁴ and cryoSPARC. Maps to fit TBP, Spt8 and the DUB module were obtained after 3D classifications using Relion 3 by applying masks covering the region of interest.

To describe the flexibility of the DUB and HAT modules, the selected SAGA images was analyzed using the Imagic software package⁶⁵. The selected images were aligned against 75 equidistant projections of the 3-D model and clustered to form 1000 2-D classes. Images from classes corresponding to two different views were grouped and aligned separately against their corresponding reference. Local classification into 50 classes was performed for each group by using an adapted mask and the best class averages were selected to show the extent of protein density variation within each view.

Model Building.

Pairwise alignment was performed with Needle⁶⁶. Homology models for Tra1, TBP, Spt8, DUB module, all 8 histone folds, Taf6 heat repeats, Taf5-WD40, Spt20-SEP domain and Taf5-NTD domains were generated using iTASSER⁶⁷ or Swiss-model and docked into the maps manually. The rigid-body docking of DUB and Spt8 was guided by the overall shape of the local density which lacked clear secondary structure features, probably due to the peripheral position and flexible nature of their connection to SAGA core. The position of the DUB is consistent with previous studies, notably XL/MS¹⁴ and negative

stained particles with GFP labeled subunits of the module²³. As the DUB is composed of two lobes with similar overall shape we cannot exclude at the current resolution an inverted (upside-down) orientation for the DUB. The location and orientation of Spt8 is validated by cross-linking mass spectrometry¹⁴, specific single-site photo activated cross-linking^{33,44} and biochemical data³⁹.

The other homology models, with the notable exception of TBP where only a minor tilt of the C-terminal stirrup was applied, were re-built according to density using Coot⁶⁸ and then extended of their N' and C' termini until a significant discontinuity in the map was encountered. Such discontinuities normally corresponded to sites where RaptorX⁶⁹ and PSIPRED⁷⁰ predictions indicated a large disordered region. To assign densities that were not clearly connected to any of these extended starting models we considered the local map properties including secondary structure features, proteins in the vicinity, biochemical data, cross-linking information for surrounding domains and most importantly the fit between protein side chains and the density. These parts of the structure were then built de-novo. For a large proportion of the side chains density was clear and practically all densities in the core of SAGA and Tra1 could be assigned with confidence. The single exception to this procedure is the Ada3 domain at the docking site of the HAT module where density showed secondary structure elements but could not resolve individual side chains. We cannot therefore exclude errors in amino acids register for this domain. Though domain deletion analysis suggested that the SWIRM domain at the C-terminus of Ada2 is also involved in docking the HAT module, we did not find densities that could correspond to this bulky domain. Tra1 and the main lobe were separately geometry-optimized in Phenix⁷¹ followed by real-space refinement with secondary structure restrains against the corresponding post-processed maps. This procedure was repeated after a round of final manual corrections to the structure in Coot. All display images were generated using UCSF Chimera⁷² and ChimeraX⁷³.

Data Availability.

The cryo-EM maps have been deposited in the 3D-EM database (EMBL-European Bioinformatics Institute, Cambridge, UK). EMBD accession codes are XXXX. The model coordinates were deposited in the PDB database accession code YYYY.

References and Notes

- 1 Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends.Biochem.Sci.* **21**, 327-335 (1996).
- 2 Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).
- 3 Nogales, E., Louder, R. K. & He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **46**, 59-83 (2017).
- 4 Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-143 (2015).
- 5 Hahn, S. & Young, E. T. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705-736 (2011).
- 6 Larschan, E. & Winston, F. The *S. cerevisiae* SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. *Genes Dev* **15**, 1946-1956 (2001).
- 7 Bhaumik, S. R. & Green, M. R. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters in vivo. *Mol Cell Biol* **22**, 7365-7371 (2002).
- 8 Baptista, T. *et al.* SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Mol Cell* **68**, 130-143 e135 (2017).
- 9 Warfield, L. *et al.* Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. *Mol Cell* **68**, 118-129 e115 (2017).
- 10 Lee, K. K. *et al.* Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Mol Syst Biol* **7**, 503 (2011).
- 11 Grant, P. A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**, 1640-1650 (1997).
- 12 Henry, K. W. *et al.* Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**, 2648-2663 (2003).
- 13 Brown, C. E. *et al.* Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* **292**, 2333-2337. (2001).
- 14 Han, Y., Luo, J., Ranish, J. & Hahn, S. Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J* **33**, 2534-2546 (2014).
- 15 Kohler, A., Zimmerman, E., Schneider, M., Hurt, E. & Zheng, N. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* **141**, 606-617 (2010).
- 16 Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**, 725-728 (2016).
- 17 Samara, N. L. *et al.* Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* **328**, 1025-1029 (2010).
- 18 Rojas, J. R. *et al.* Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* **401**, 93-98 (1999).
- 19 Sun, J. *et al.* Structural basis for activation of SAGA histone acetyltransferase Gcn5 by partner subunit Ada2. *Proc Natl Acad Sci U S A* **115**, 10010-10015 (2018).
- 20 Diaz-Santin, L. M., Lukoyanova, N., Aciyan, E. & Cheung, A. C. Cryo-EM structure of the SAGA and NuA4 coactivator subunit Tra1 at 3.7 angstrom resolution. *eLife* **6** (2017).

- 21 Sharov, G. *et al.* Structure of the transcription activator target Tra1 within the chromatin modifying complex SAGA. *Nat Commun* **8**, 1556 (2017).
- 22 Wu, P. Y., Ruhlmann, C., Winston, F. & Schultz, P. Molecular architecture of the *S. cerevisiae* SAGA complex. *Mol Cell* **15**, 199-208 (2004).
- 23 Setiaputra, D. *et al.* Conformational flexibility and subunit arrangement of the modular yeast spt-ada-gcn5 acetyltransferase complex. *J Biol Chem* **290**, 10057-10070, (2015).
- 24 Durand, A., Bonnet, J., Fournier, M., Chavant, V. & Schultz, P. Mapping the deubiquitination module within the SAGA complex. *Structure* **22**, 1553-1559 (2014).
- 25 Grant, P. A. *et al.* A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* **94**, 45-53 (1998).
- 26 Ogryzko, V. V. *et al.* Histone-like TAFs within the PCAF histone acetylase complex *Cell* **94**, 35-44 (1998).
- 27 Gangloff, Y., Romier, C., Thuault, S., Wertén, S. & Davidson, I. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci* **26**, 250-257. (2001).
- 28 Kolesnikova, O. *et al.* Molecular structure of promoter-bound yeast TFIID. *Nature communications* **9**, 4666 (2018).
- 29 Patel, A. B. *et al.* Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362** (2018).
- 30 Eisenmann, D. M., Arndt, K. M., Ricupero, S. L., Rooney, J. W. & Winston, F. SPT3 interacts with TFIID to allow normal transcription in *Saccharomyces cerevisiae*. *Genes Dev.* **6**, 1319-1331 (1992).
- 31 Laprade, L., Rose, D. & Winston, F. Characterization of new Spt3 and TATA-binding protein mutants of *Saccharomyces cerevisiae*: Spt3 TBP allele-specific interactions and bypass of Spt8. *Genetics* **177**, 2007-2017 (2007).
- 32 Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev* **22**, 2994-3006 (2008).
- 33 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).
- 34 Hoffmann, A. *et al.* A histone octamer-like structure within TFIID. *Nature* **380**, 356-359. (1996).
- 35 Birck, C. *et al.* Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* **94**, 239-249 (1998).
- 36 Selleck, W. *et al.* A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. *Nat Struct Biol* **8**, 695-700. (2001).
- 37 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
- 38 Nikolov, D. B. *et al.* Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119-128 (1995).
- 39 Wu, P. Y. & Winston, F. Analysis of Spt7 function in the *Saccharomyces cerevisiae* SAGA coactivator complex. *Mol Cell Biol* **22**, 5367-5379 (2002).
- 40 Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830-836 (1996).
- 41 Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127-151 (1996).
- 42 Wollmann, P. *et al.* Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. *Nature* **475**, 403-407 (2011).
- 43 Imbalzano, A. N., Zaret, K. S. & Kingston, R. E. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol Chem.* **269**, 8280-8286 (1994).
- 44 Warfield, L., Ranish, J. A. & Hahn, S. Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes Dev* **18**, 1022-1034 (2004).
- 45 Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* **8** (2019).

- 46 Liu, D. *et al.* Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**, 573-583 (1998).
- 47 Anandapadamanaban, M. *et al.* High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat Struct Mol Biol* **20**, 1008-1014 (2013).
- 48 Kamata, K. *et al.* C-terminus of the Sgf73 subunit of SAGA and SLIK is important for retention in the larger complex and for heterochromatin boundary function. *Genes Cells* **18**, 823-837 (2013).
- 49 Elias-Villalobos, E., Toullec, D., Faux, C., Seveno, M. & helmlinger, D. Chaperone-mediated ordered assembly of the SAGA and NuA4 transcription co-activator complexes. *bioRxiv* **524959** (2019).
- 50 Knutson, B. A. & Hahn, S. Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol Cell Biol* **31**, 818-831 (2011).
- 51 Mutiu, A. I. *et al.* Structure/function analysis of the phosphatidylinositol-3-kinase domain of yeast tra1. *Genetics* **177**, 151-166 (2007).
- 52 Mittal, C., Culbertson, S. J. & Shogren-Knaak, M. A. Distinct requirements of linker DNA and transcriptional activators in promoting SAGA-mediated nucleosome acetylation. *J Biol Chem* **293**, 13736-13749 (2018).
- 53 Saint, M. *et al.* The TAF9 C-terminal conserved region domain is required for SAGA and TFIID promoter occupancy to promote transcriptional activation. *Mol Cell Biol* **34**, 1547-1563 (2014).
- 54 Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol Cell* **13**, 573-585 (2004).
- 55 Ravarani, C. N., Chalancon, G., Breker, M., de Groot, N. S. & Babu, M. M. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat Commun* **7**, 10417 (2016).
- 56 Gupta, K. *et al.* Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *eLife* **6** (2017).
- 57 Ranish, J. A., Lane, W. S. & Hahn, S. Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science* **255**, 1127-1129 (1992).
- 58 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
- 59 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
- 60 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
- 61 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7** (2018).
- 62 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).
- 63 Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7** (2018).
- 64 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65 (2014).
- 65 van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17-24 (1996).
- 66 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641 (2019).
- 67 Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174-181 (2015).
- 68 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
- 69 Kallberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7**, 1511-1522 (2012).
- 70 Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, W349-357 (2013).
- 71 Terwilliger, T. C. Rapid model building of alpha-helices in electron-density maps. *Acta Crystallogr D Biol Crystallogr* **66**, 268-275 (2010).

- 72 Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J Struct Biol* **157**, 281-287 (2007).
- 73 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).
- 74 Zhou, Q. & Berk, A. J. The yeast TATA-binding protein (TBP) core domain assembles with human TBP-associated factors into a functional TFIID complex. *Mol Cell Biol* **15**, 534-539 (1995).
- 75 Kelleher, R. J. *et al.* Yeast and human TFIIDs are interchangeable for the response to acidic transcriptional activators in vitro. *Genes.Dev.* **6**, 296-303 (1992).
- 76 Ozer, J., Lezina, L. E., Ewing, J., Audi, S. & Lieberman, P. M. Association of transcription factor IIA with TATA binding protein is required for transcriptional activation of a subset of promoters and cell cycle progression in *Saccharomyces cerevisiae*. *Mol Cell Biol* **18**, 2559-2570 (1998).

Acknowledgements

We thank Didier Devys and Laszlo Tora for carefully reading the manuscript and for advises. We are indebted to Renaud Wagner for advice during the initial stages of this project. We are grateful to Robin Ben-Shem for graphics expertise. We acknowledge support from the Institut National de la Santé et de la Recherche Médicale (INSERM), the Centre National pour la Recherche Scientifique (CNRS), the Association pour la Recherche sur le Cancer (ARC), the Ligue contre le Cancer, the ANR-15-CE11-0022-01 and the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d’Avenir ANR-10-IDEX-0002-02. We acknowledge the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC.

Author contributions.

P.S. and A.B.S. designed the study; A.B.S originated SAGA purification method and reconstituted the SAGA-TBP complex; A.B.S. and C.C. defined conditions for grid preparation and freezing ; C.C. and G.P. froze grids ; G.P. collected and analyzed cryo-EM data; A.F., G.P. and A.B.S. interpreted the maps by fitting crystal coordinates and model building ; O.K. purified TBP, TFIIA and SAGA and used them to perform pull-down and gel-shift assays; P.S. and A.B.S. supervised the work; G.P. and O.K. prepared figures ; A.B.S and P.S. wrote the manuscript with input from all authors.

Author information.

The authors declare no competing financial interest. Correspondence and request for materials should be addressed to P.S. (patrick.schultz@igbmc.fr) or A.B.S. (adam@igbmc.fr).

Figures

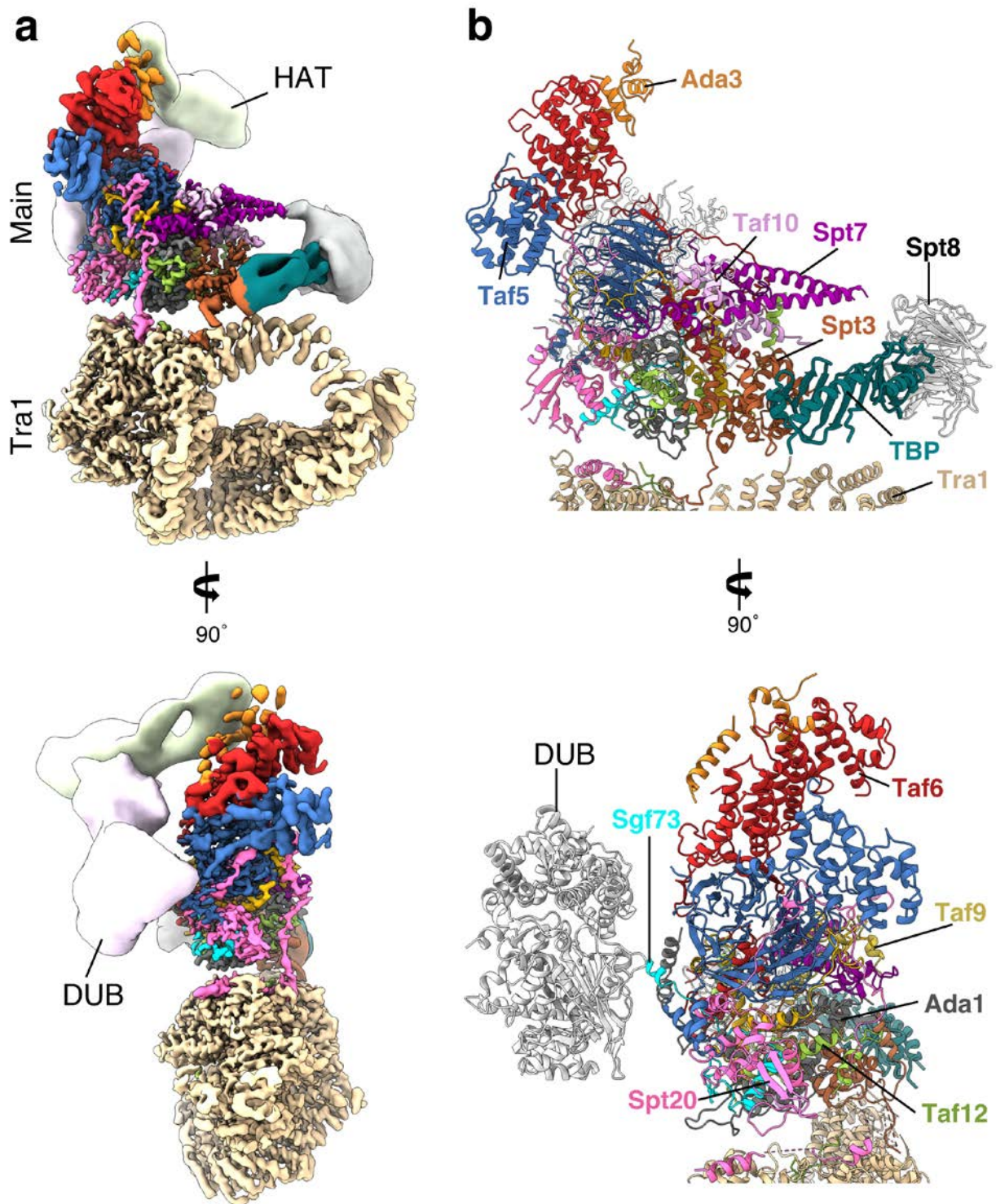


Figure 1 | Structure of the yeast SAGA-TBP complex. **a**, Front and side views of a composite cryo-EM reconstruction of the yeast SAGA-TBP complex. Maps from focused refinements of Tra1 (tan), DUB (light pink), Spt8 (white), the main lobe and TBP (forest green) were combined. HAT density (light green) taken from low-pass filtered map of the full complex. **b**, Corresponding views of the atomic model of the SAGA-TBP complex focusing on the main lobe and showing the central module (subunits Spt7, Spt3, Ada1, Spt20 and the five TAFs), Spt8, TBP, the DUB module, Sgf73 and Ada3.

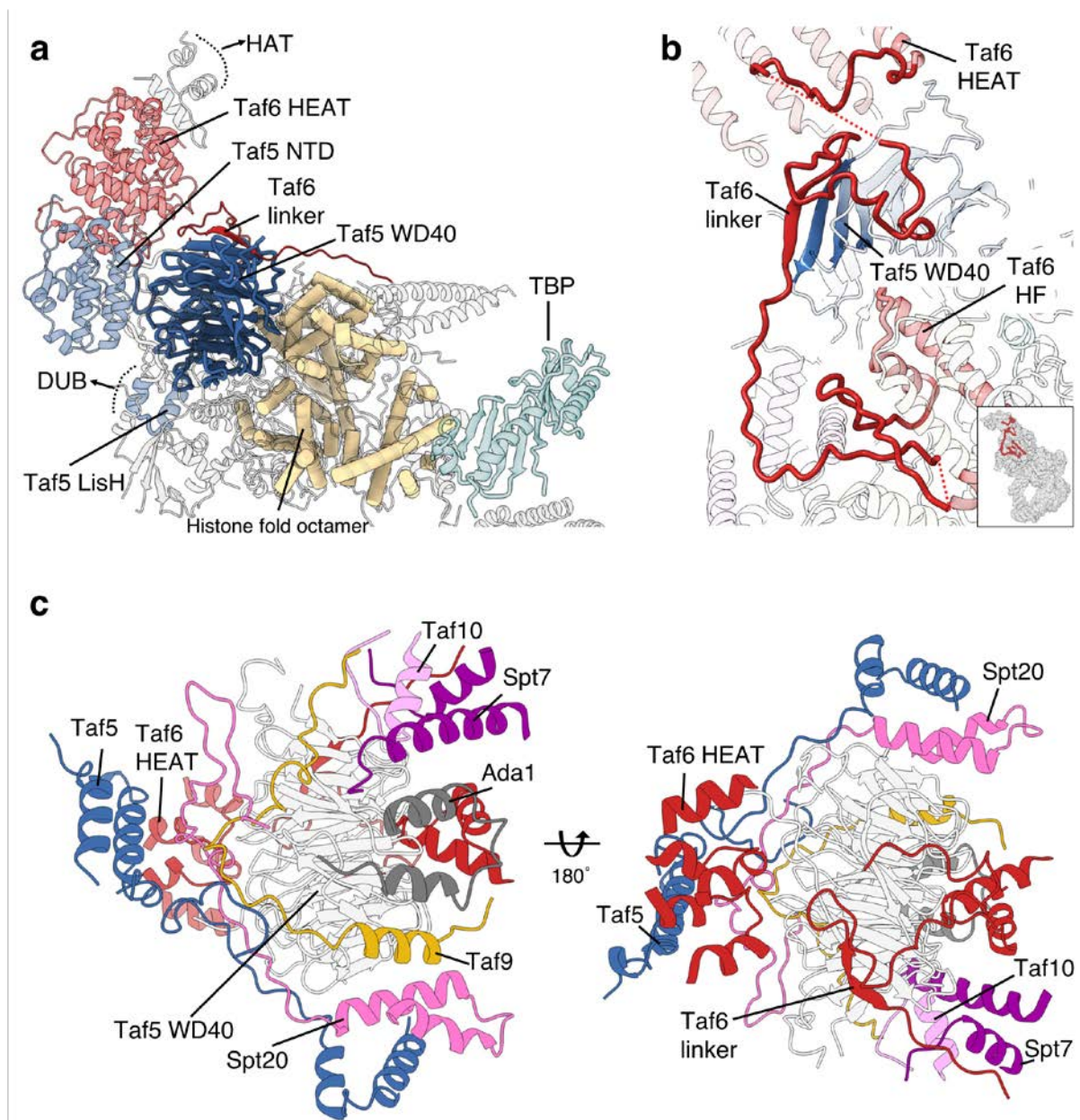


Figure 2 | Taf5 structurally orchestrates the main lobe. **a**, Central position of the Taf5 WD40 domain, located between Taf5-NTD and Taf6-HEAT repeats on one side and the histone-folds machinery on the other. Orientation as in Fig. 1a. **b**, Linker connecting Taf6 HEAT repeats to its HF domain forming an inter-protein beta sheet with the Taf5 WD40 propeller. In the small insert the depicted region is highlighted within the overall structure of SAGA. Disordered loops residues are indicated with dotted lines. **c**, Intricate network of protein contacts with the Taf5 WD40 domain (white).

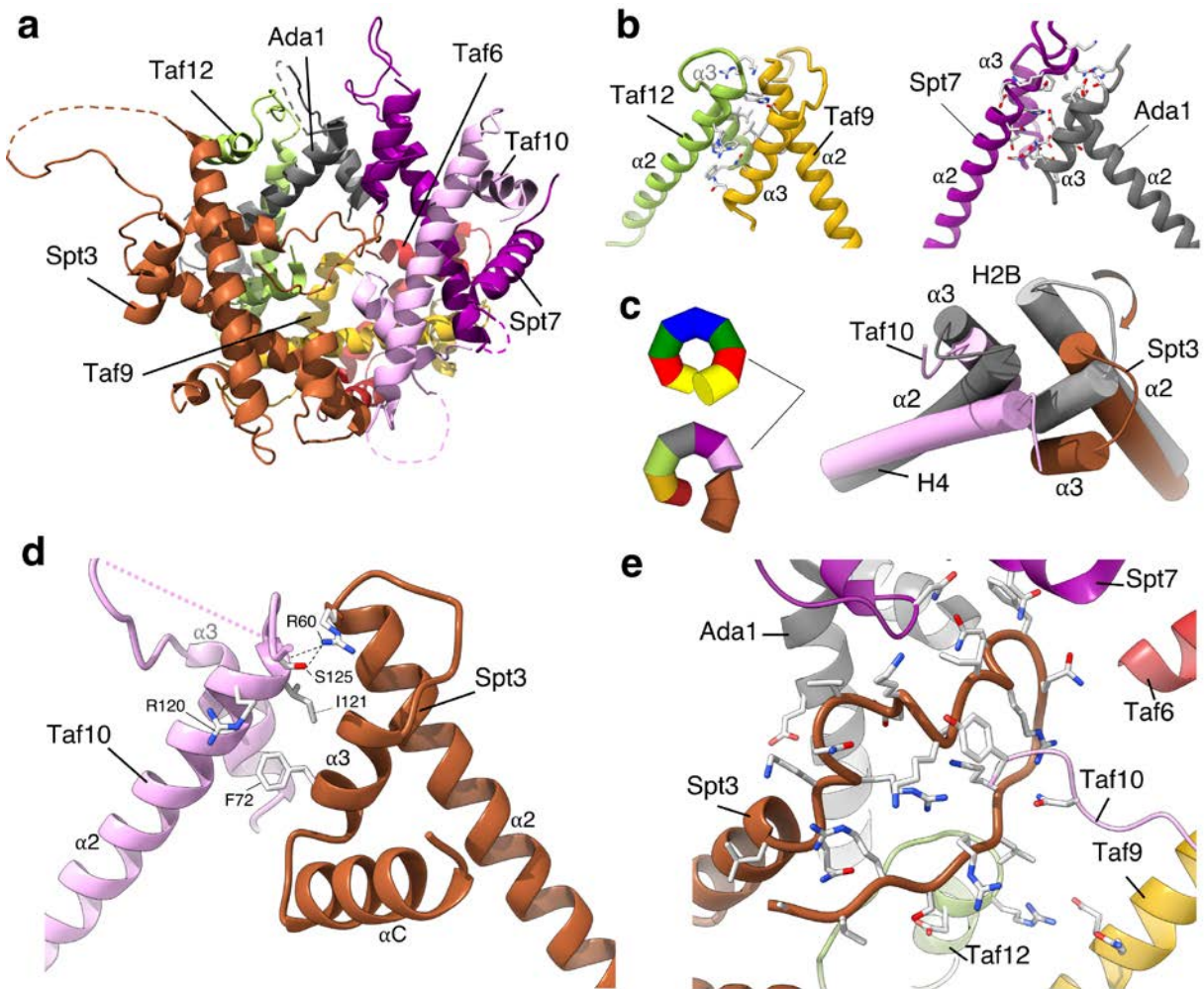


Figure 3 | The octamer of histone folds. **a**, Structural organization of the four pairs of histone folds (Taf6-Taf9, Ada1-Taf12, Spt7-Taf10 and nSpt3-cSpt3) forming an octamer analogous to the nucleosomal histones. **b**, Structure of the four-helix bundles connecting the Taf6-Taf9 HF pair to the Ada1-Taf12 HF pair and the Ada1-Taf12 HF pair to the Spt7-Taf10 HF pair. Only side chains of residues that participate in the interactions are shown. **c**, Left panel: Schematic representation of the HF octamer in the nucleosome (top) and in SAGA (bottom). Right panel: The four-helix bundle connecting the Spt7-Taf10 HF pair (pink) to the partially dislodged nSpt3-cSpt3 HF pair (brown) compared to the analogous nucleosome interaction (dark grey). **d**, Detailed view of the interaction between the Spt7-Taf10 HF pair and the nSpt3-cSpt3 HF pair. **e**, The C-terminal tail of Spt3 inserts into the center of the HF octamer and interacts with each of the other histone folds.

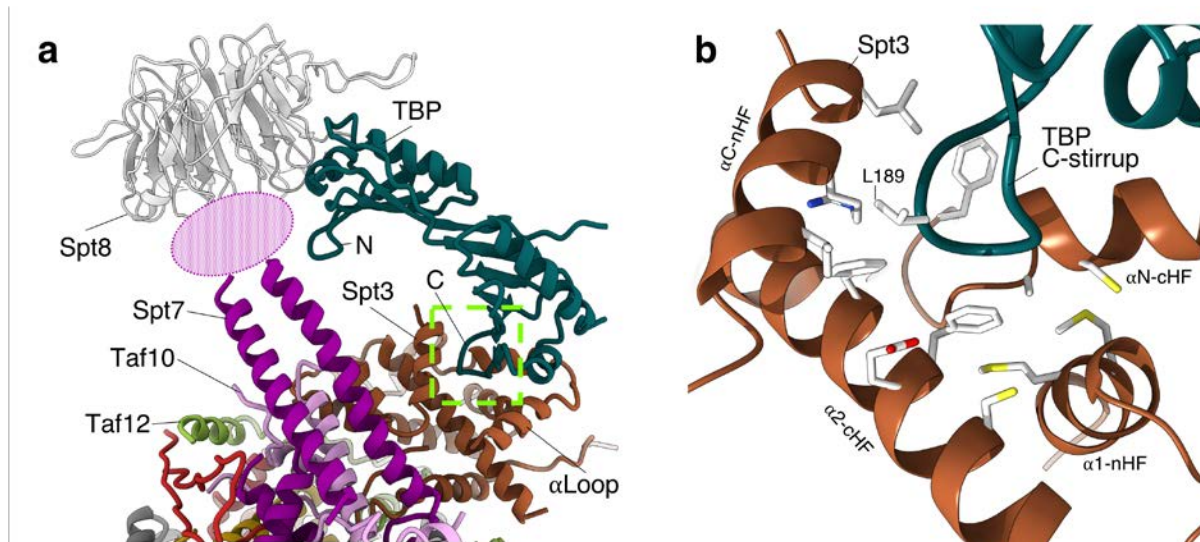


Figure 4 | Architecture of the TBP docking site. **a**, cTBP interacts with Spt3 mainly through the stirrup of its C-terminal half (labeled C) while its N-terminal half (labeled N) binds to the Spt8 WD40 domain. A poorly resolved Spt7 domain (pink ellipse) connects Spt8 to the central module. **b**, Blow-up of the boxed area in (a) depicting the interaction between the C-terminal stirrup of TBP and a pocket created by the α N helix of the C-terminal Spt3 Histone fold (α N-cHF), the α 2 helix of the C-terminal Spt3 Histone fold (α 2-cHF) and the α C helix joined to the nSpt3-HF.

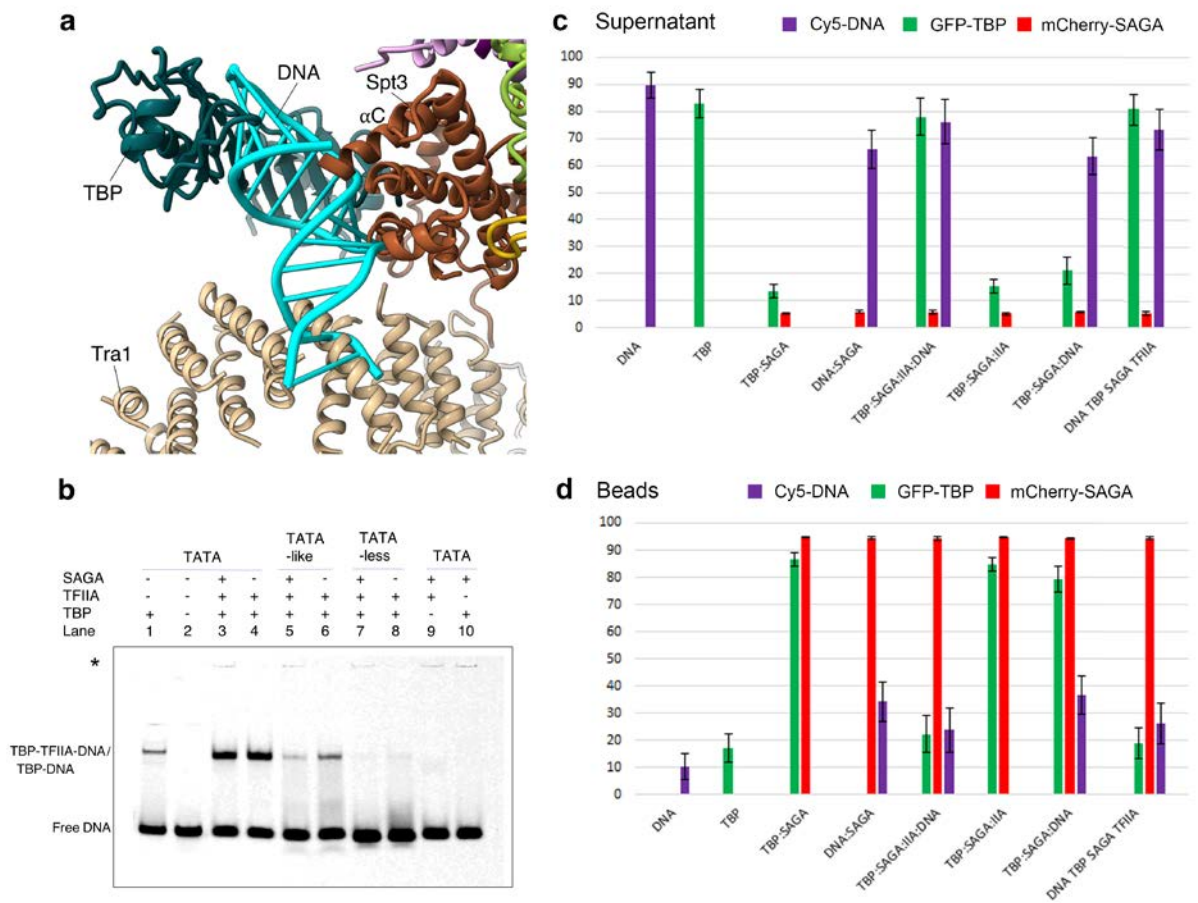


Figure 5 | Steric hindrance for TBP binding to DNA and TFIIA-dependent TBP delivery. **a**, DNA (light blue) was modeled into the DNA binding cleft of TBP by superposing the structure of cTBP-DNA (pdb 1YTF) on SAGA-bound cTBP. DNA clashes with Tra1 HEAT-repeats and with Spt3 helix α C. **b**, Gel-shift assay using a Cy5-labeled TATA-containing CYC1 promoter DNA probe (TATA), a TATA-like DNA probe and a non-specific DNA probe (TATA-less). TBP was pre-incubated with purified SAGA (devoid of TBP as shown in Extended Data Fig. 9e) to form SAGA-TBP complex before addition of DNA probes and TFIIA (as indicated). The content of shifted bands was verified by mass-spectrometry (Extended Data Fig. 9d). The asterisk indicates minor non-specific DNA association with SAGA. **c-d**, Indicated components are incubated as for the gel shift assays and SAGA is immobilized by streptavidin beads that bind SBP-tagged SAGA (Extended Data Fig. 9a). DNA was labeled by Cy5 (red), TBP by GFP (green), SAGA by mCherry (blue). Proportion of fluorescent signal due to the unbound (c) and bound (d) complexes are presented as percent from the total signal (bound and unbound). In the right-most bars TBP was first incubated with DNA and TFIIA prior to the addition of SAGA.

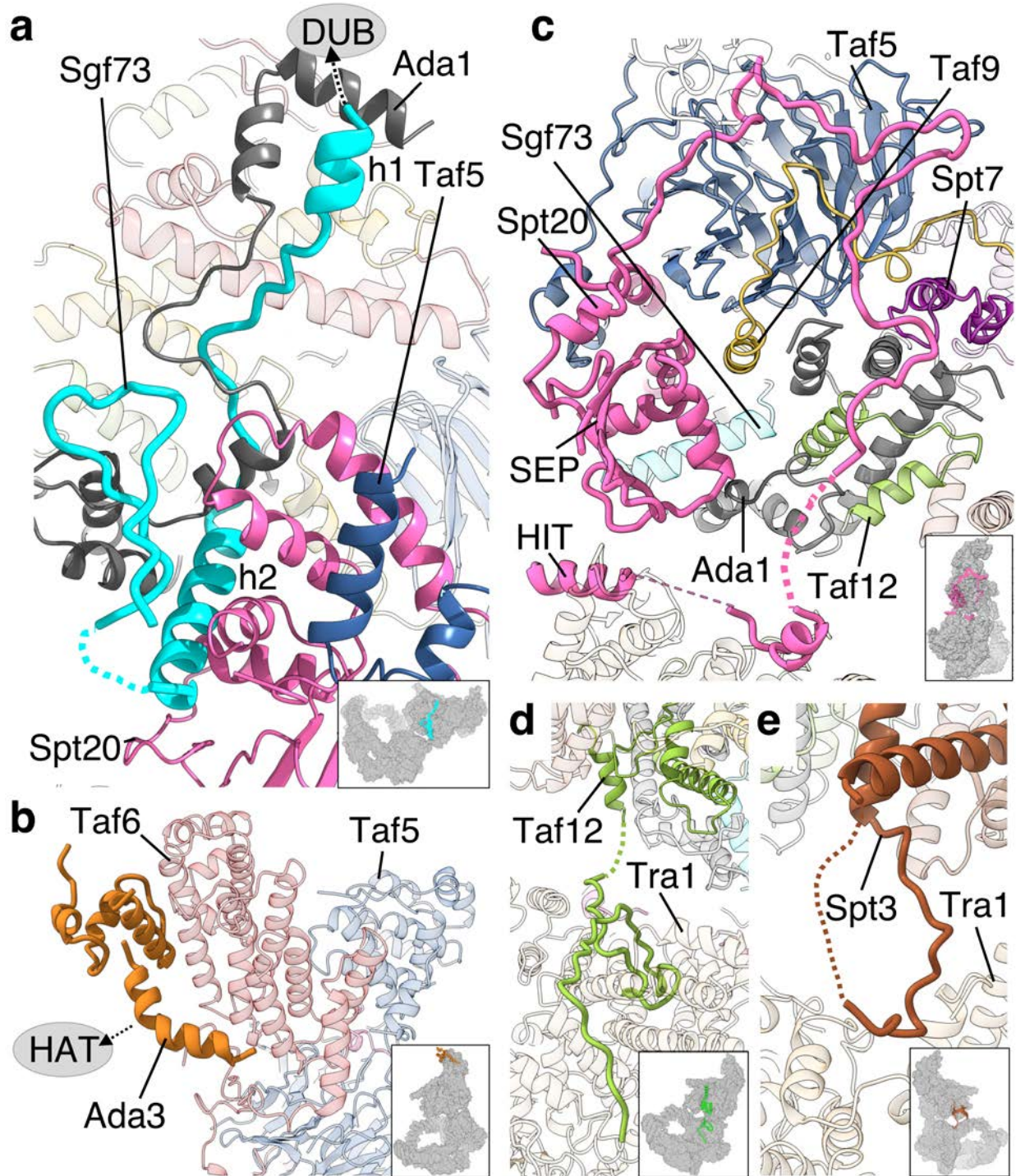
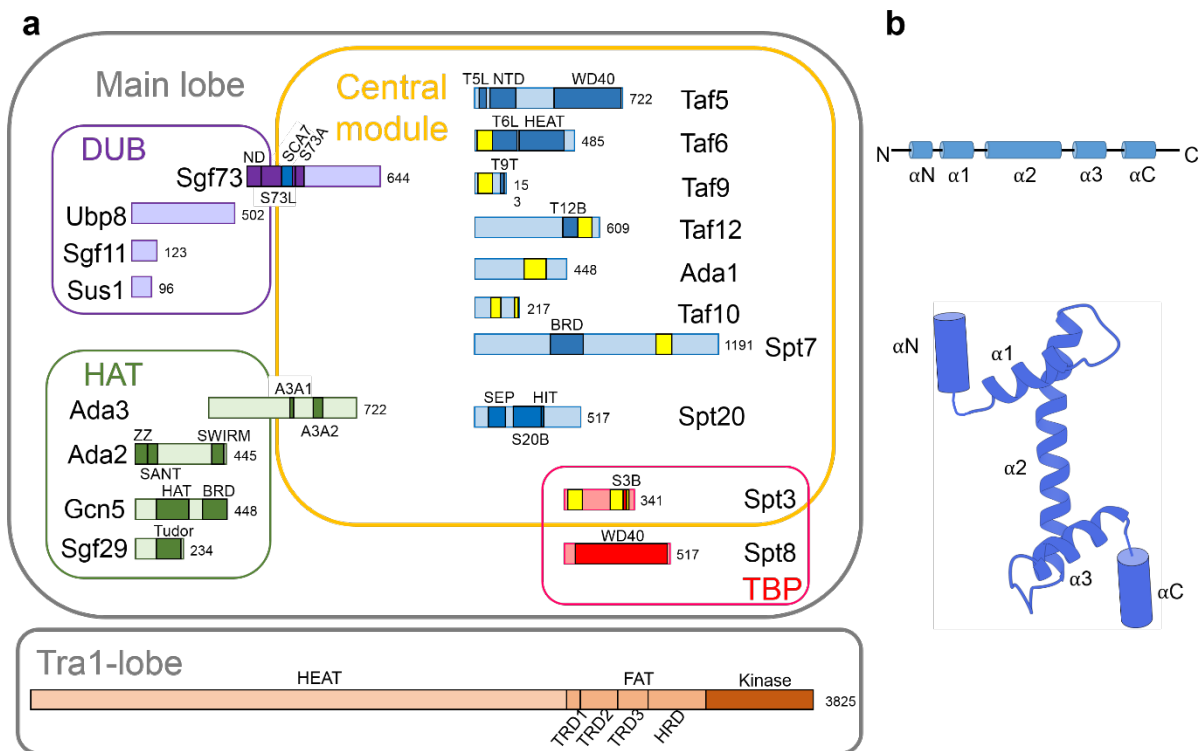


Figure 6 | Connecting the enzymatic modules and Tra1 to the central module. **a**, The Sgf73 domain that harbors helices Sgf73-h1 and Sgf73-h2 (cyan) anchors the DUB module within the central module. Sgf73-h2 is embedded in a chamber formed by Spt20 (pink) and Ada1 (dark grey). **b**, Two helical domains of Ada3 (ochre) anchor the HAT module at the surface of the Taf6 HEAT repeats (red). **c-e**, Bridges between the central module and Tra1 (tan) are established by **(c)** Spt20 (pink) **(d)** Taf12 (green) and **(e)** the loop connecting the second and third helices of cSpt3-HF (brown).

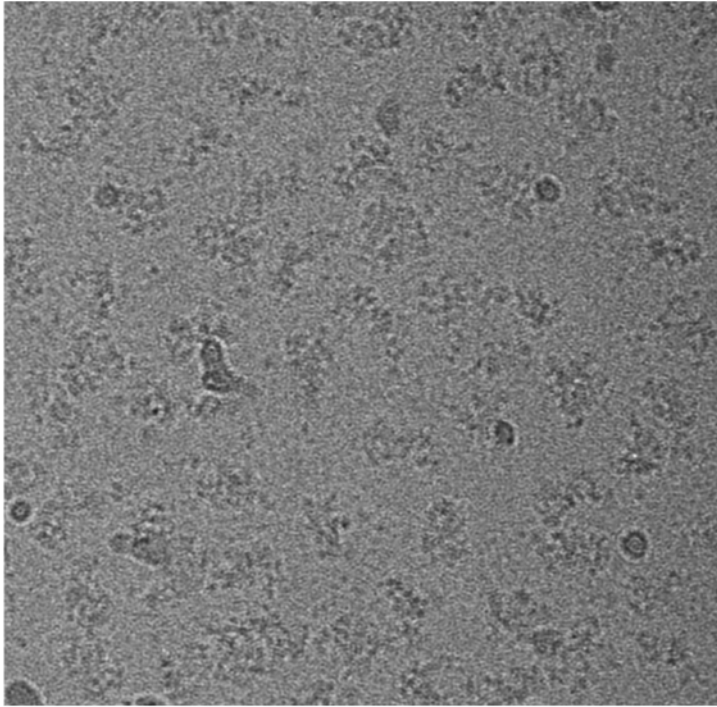
Extended data



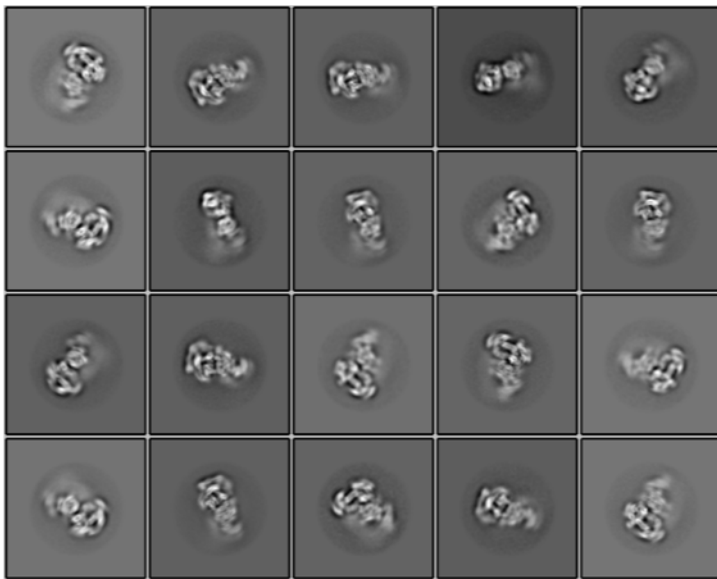
Extended Data Figure 1 | Subunit and domain organization of the yeast SAGA complex

a, SAGA subunits are organized into the Tra1-lobe and the main lobe. The main lobe is divided into two enzymatic HAT and DUB modules, a central module and a TBP-binding module. Overlaps appear as the central module is a scaffold for the assembly of all other modules. Abbreviations used due to space limitations: Taf5 Lissencephaly Homology (LisH) domain (**T5L**), Taf6 linker domain (N83-S218, **T6L**), Taf9 C-tail (N110-L139, **T9T**), Sgf73 N-terminal domain (M1-K69, **ND**), Sgf73 linker region (I70-H234, **S73L**), Sgf73 anchoring domain (E235-L277, **S73A**), Spt20 HIT domain interacting with Tra1 (M325-L340, **HIT**), Ada3 anchoring domains (I396-G415 **A3A1** and Y509-E557 **A3A2**), bromodomain (**BRD**). The **SEP** domain found in Spt20 is named after *S. cerevisiae* Shp1, *Drosophila melanogaster* eyes closed gene (*Eyc*), and vertebrate P47. The FAT (**FRAP**, **ATM** and **TRRAP**) repeats are divided into three TRD (Tetratricopeptide Repeat-containing Domain) domains and one HRD (**HEAT Repeat Domain**) domain. We defined three bridge forming domains involved in connecting Tra1 to the central module: the Spt20 bridge (K184-K336, **S20B**), the Taf12 bridge (T430-K502, **T12B**) and the Spt3 bridge (E268-I304, **S3B**). The Histone Fold domains, present in seven subunits, are highlighted as yellow boxes. **b**, Schematic representation and helix nomenclature for histone fold domains.

a

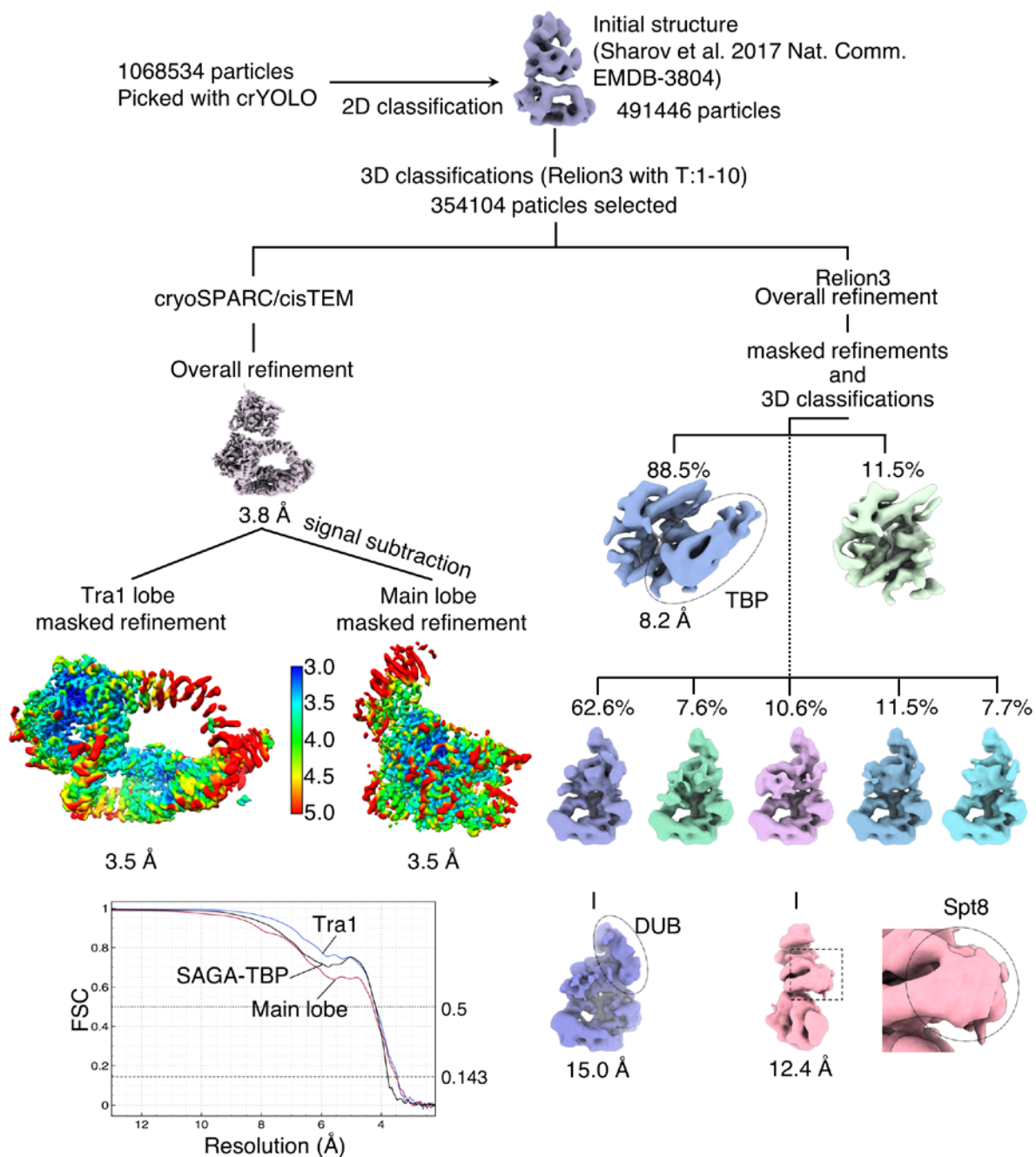


b



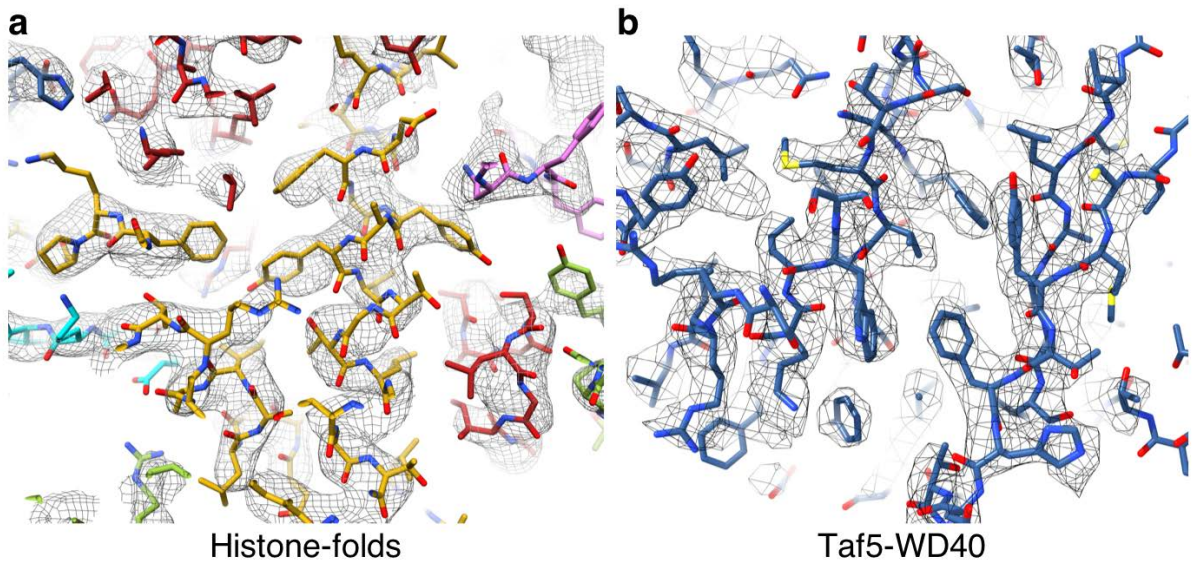
Extended Data Figure 2 | Single-particle cryo-EM analysis of the SAGA-TBP complex

a, Original image of frozen-hydrated *P. pastoris* SAGA-TBP complex. **b**, Two-dimensional classes averages showing high resolution structural features.



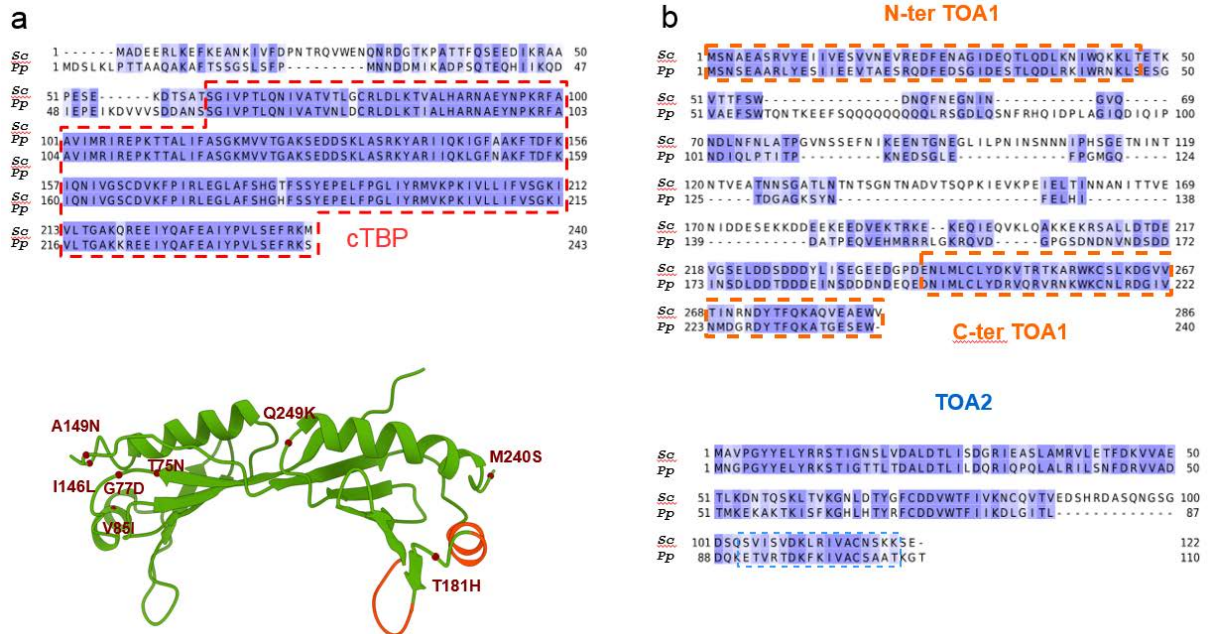
Extended Data Figure 3 | Cryo-EM data analysis strategy and resolution assessment of the cryo-EM structures

Image processing strategy used to obtain the 3.5 Å resolution maps of the SAGA-TBP complex. Maps colored in rainbow spectrum represent the local resolution of the reconstructions. Fourier Shell Correlation (FSC) curves are depicted as a function of resolution in Å for the entire SAGA-TBP complex (black), the Tra1-lobe (blue) and the Main-lobe (red). cryoSPARC v2 and cisTEM were used to generate the 3.5 Å resolution maps of the Tra1 and Main lobes. Relion 3 was used for masked refinements and classifications of flexible regions, corresponding to TBP, the DUB module and Spt8. Dotted circles show the regions, which were used to dock the crystal structures of TBP (pdb 1TBP), Spt8 (homology model by Swiss-Model using pdb 4J87 as template), and the DUB module (pdb 6AQR).



Extended Data Figure 4 | High resolution structure of the yeast SAGA complex.

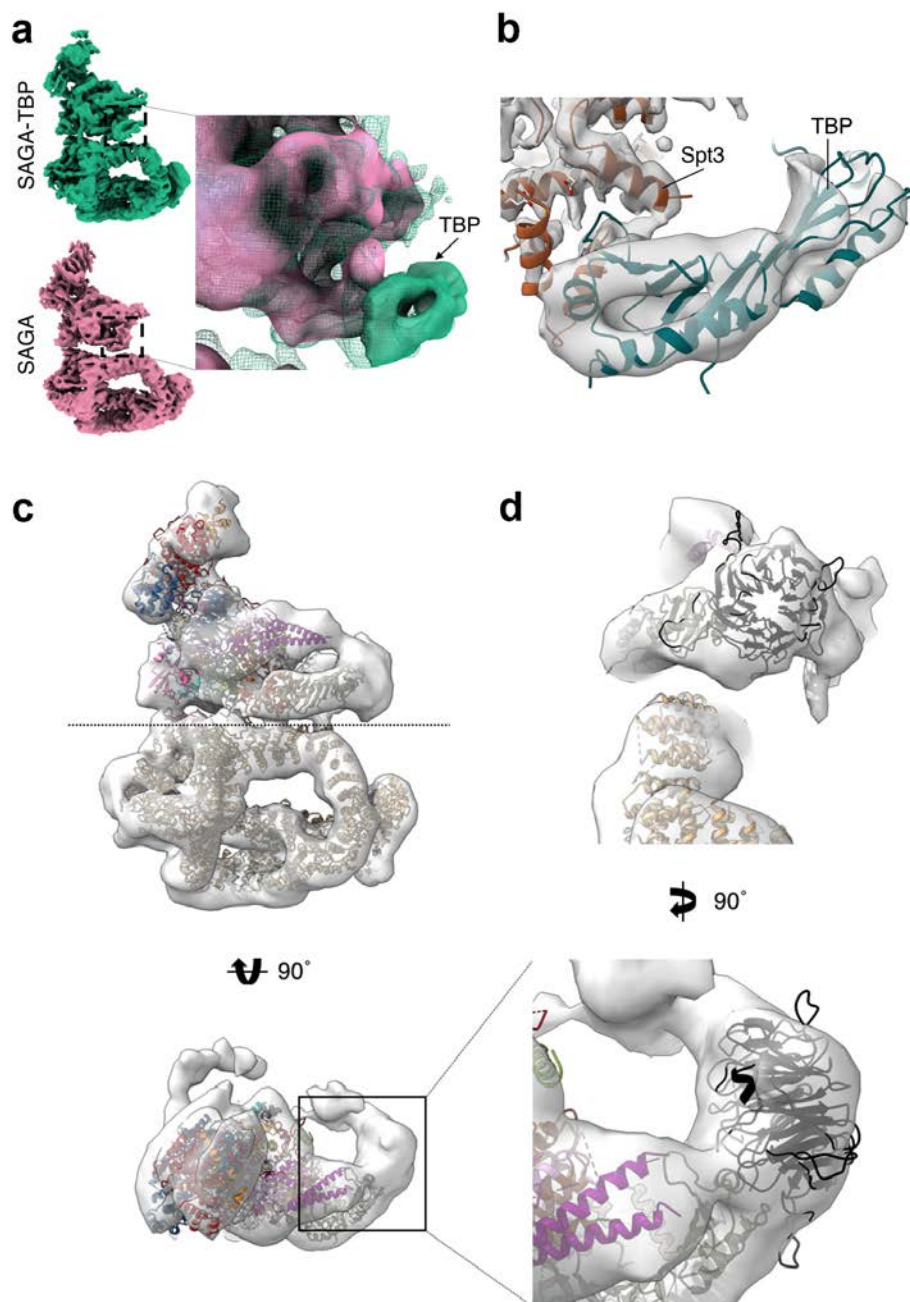
Representative regions illustrating the quality of the cryo-EM map and the high resolution structural features. Cryo-EM map and atomic model showing that side chains are clearly identified. **a**, Histone fold domain of Taf9. **b**, Part of the Taf5 WD40 domain.



Extended Data Figure 5 | *P. pastoris* cTBP and TFIIA are nearly identical to their *S. cerevisiae* homologues.

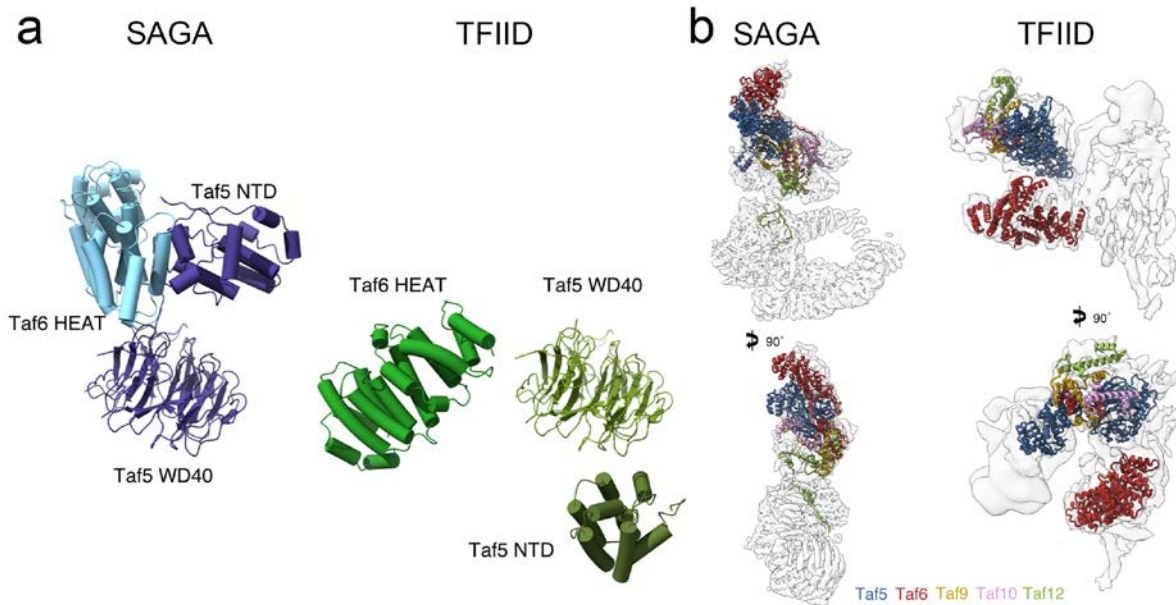
a, Pairwise alignment of TBP from *P. pastoris* and *S. cerevisiae*. Eight residues (out of 180) differ between the two organisms within the conserved C-terminal part of TBP (marked by dotted line). These residues are mapped on the structure of cTBP (pdb 1YTF). None of these residues occur in regions that contact SAGA (highlighted in red). **b**, Pairwise alignment of TFIIA subunits, Toa1 and Toa2, from *P. pastoris* and *S. cerevisiae*. The evolutionary conserved and functionally essential domains of TFIIA (marked by dotted lines) show 60% identity and 80% similarity between the two organisms.

It is worth noting that TBP and TFIIA have both been shown to be highly similar even across large evolutionary distances. For example, TBP and TOA2 from yeast and human are functionally interchangeable in Pol II transcription⁷⁴⁻⁷⁶ and yeast TFIIA complements a mammalian in-vitro transcription system depleted of TFIIA⁵⁷. Hence TBP and TFIIA from *S. cerevisiae*, which are easier to over-produce in *E. coli*, are valid substitutes in our experimental system for their homologues from the closely related budding yeast *P. pastoris*.



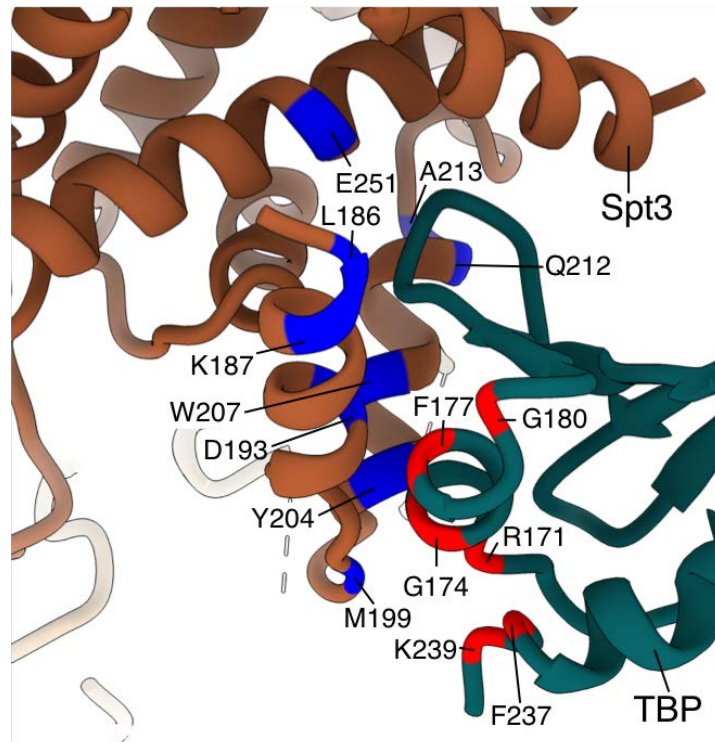
Extended Data Figure 6 | Fitting of TBP and Spt8 into the cryo-EM map.

a, Cryo-EM reconstructions determined in the presence (green) and in the absence (pink) of TBP. The enlarged panel shows the superimposition of both maps and highlights the additional density corresponding to TBP. **b**, Fitting of the cTBP crystal structure into the additional density observed in the SAGA-TBP complex. **c - d**, Fitting of the WD40 repeat of subunit Spt8 into the cryo-EM map of the SAGA-TBP complex.

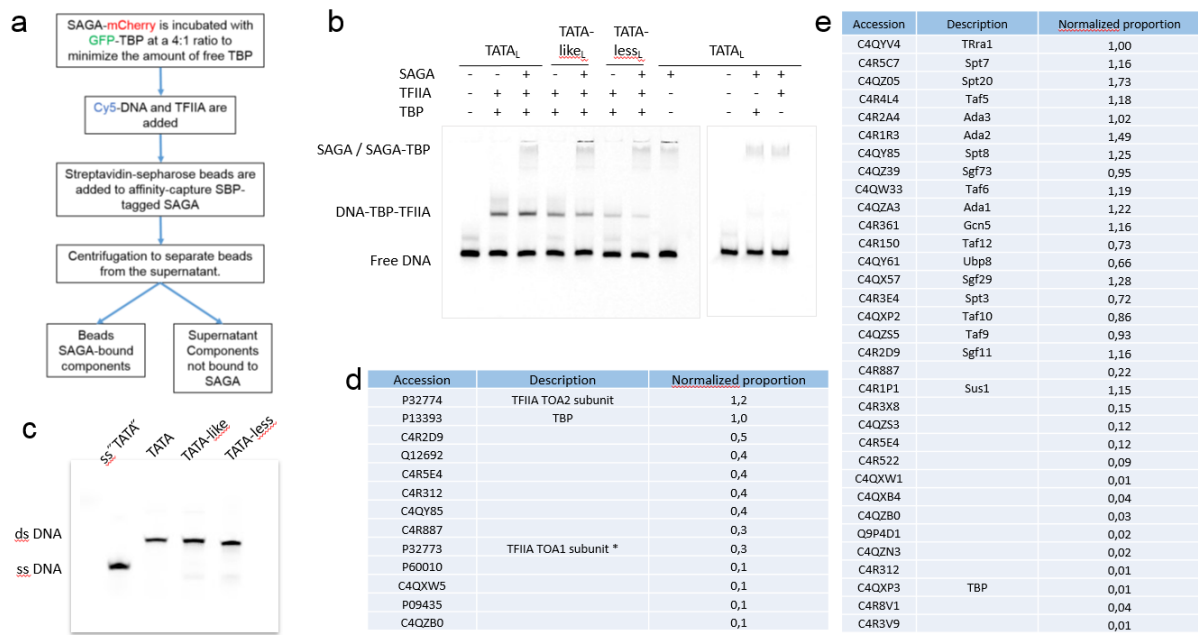


Extended Data Figure 7 | Common SAGA and TFIID subunits

a, The Taf5-NTD and Taf6 HEAT repeats have different positions in SAGA and in TFIID. The structures of Taf5-NTD, Taf6 HEAT repeats and Taf5 WD40 domain in SAGA (right) and in the twin lobe of TFIID²⁸ (left) are depicted. For the sake of clarity all other parts of SAGA and TFIID were removed. To compare the position of Taf5-NTD and Taf6 HEAT repeats between SAGA and TFIID, both complexes were aligned to have the same orientation of the Taf5 WD40 domain. The top face of the WD40 repeat, which is to a large extent free in TFIID, is nearly completely engaged in SAGA. The positions of Taf5-NTD and Taf6 HEAT repeats are defined by direct interactions of the two domains between themselves, with the WD40 surface, and with the alpha helix A352-D364 of Taf5. **b**, SAGA and TFIID share five subunits (Taf5 and the four Histone Folds-containing Taf6, Taf9, Taf10 and Taf12 subunits). This subunits are present as single or double copies in SAGA and TFIID, respectively. The relative positions of these subunits are represented in both complexes.

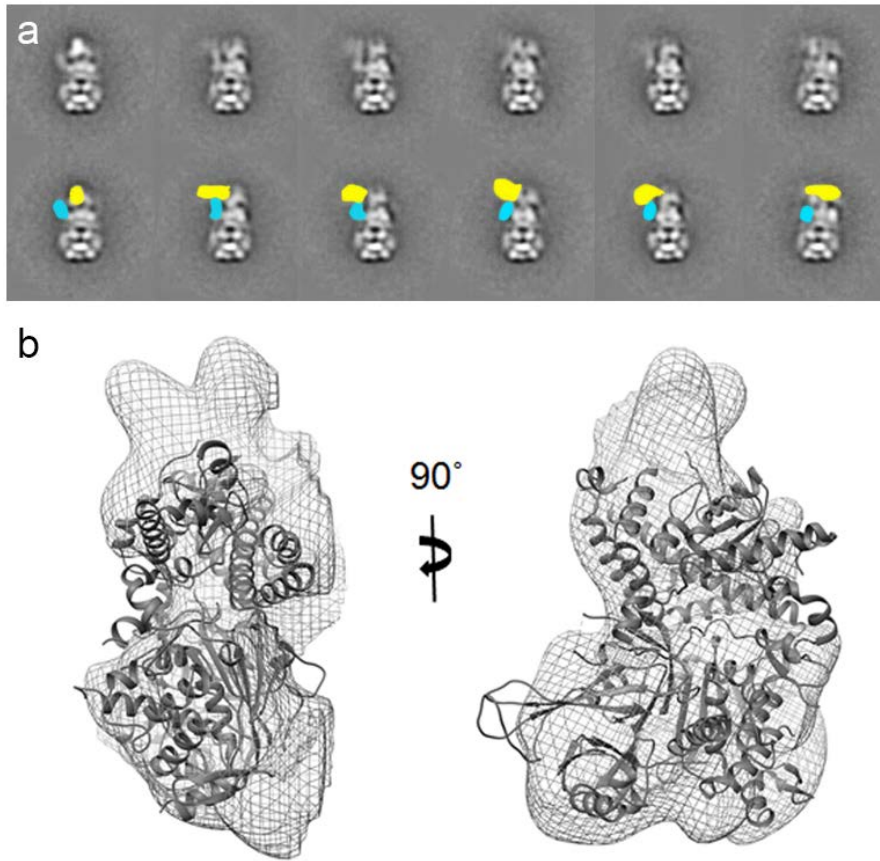


Extended Data Figure 8 | Mapping previously isolated mutations in Spt3 and TBP. Several classes of Spt3 and TBP mutations were shown to interact genetically with respect to transcription. Mutations from two major studies^{30,31} are mapped on the structure of Spt3 and TBP in the SAGA-TBP complex. The mutations appear to cluster at the contact sites between TBP and Spt3.



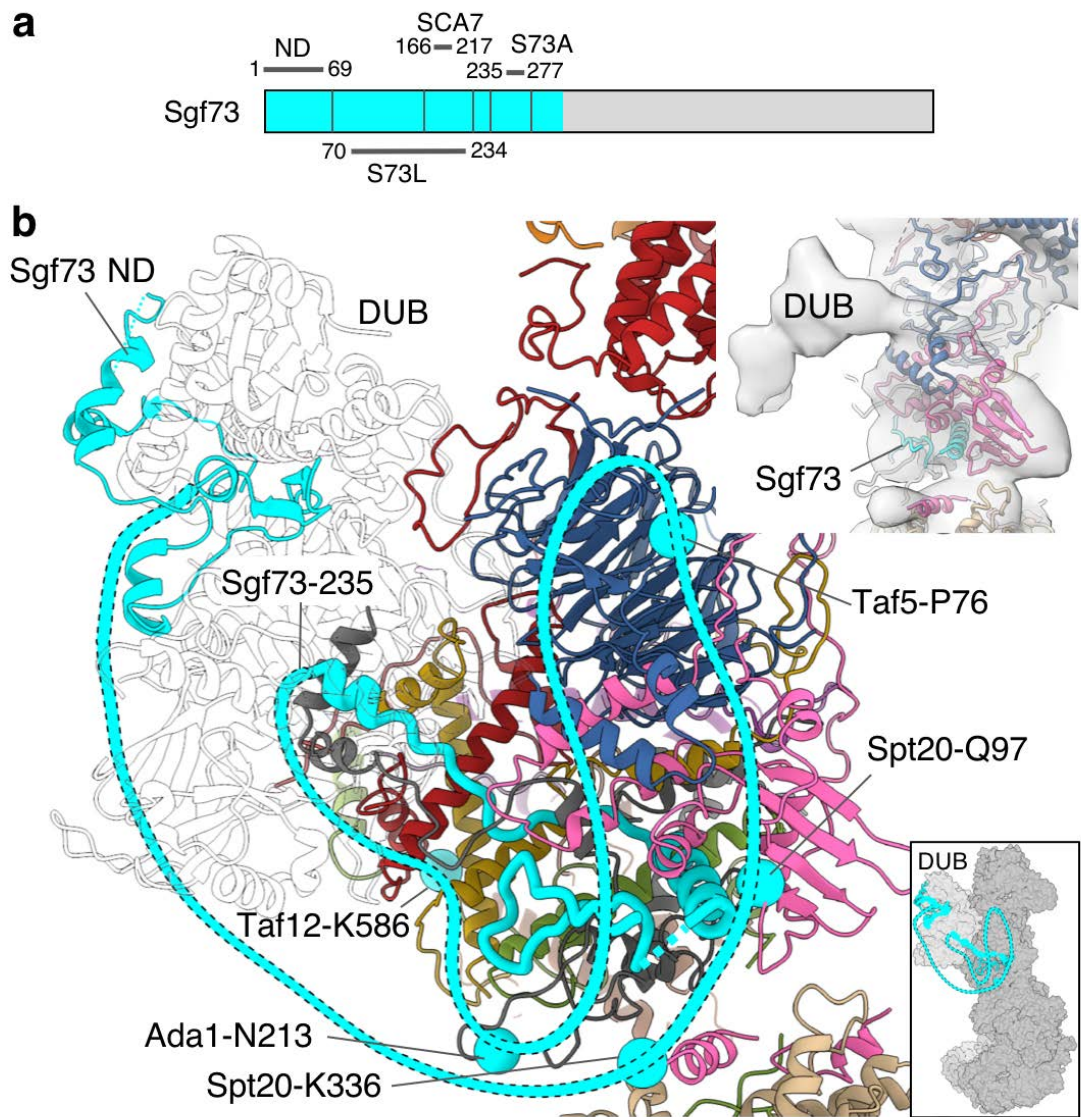
Extended Data Figure 9 | TFIIA-dependent TBP release from SAGA.

a, Experimental setup for pull-down assays that investigate TBP and DNA interactions with SBP-tagged SAGA. **b**, Gel-shift assay as presented in Fig 5b. but employing Cy5-labeled long (100 bp) DNA : TATA-containing promoter (TATA_L), a TATA-like DNA probe (TATA-like_L) and a non-specific DNA probe (TATA-less_L) **c**, Annealed Cy5-labeled DNA used in gel-shift and pull-down assays was loaded on a polyacrylamide 12% gel. Single stranded Cy5-labeled TATA-DNA was loaded for comparison. **d**, Summary of mass-spectrometry data for the major shifted DNA band (Fig. 5, lane 3) showing that it contains both TBP and TFIIA as its major constituents. PSM/AA of each protein were normalized with respect to PSM/AA of TBP. **e**, Summary of mass-spectrometry data of purified SAGA showing that TBP is present in only about 1% of the purified complexes. To estimate stoichiometry, PSM divided by the amino acid length (PSM/AA) was calculated for each protein and normalized with respect to PSM/AA of subunit Tra1. The raw data is provided as Extended Data 1.



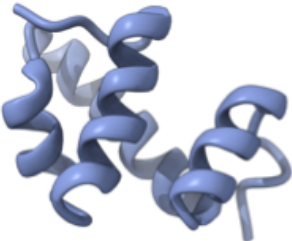




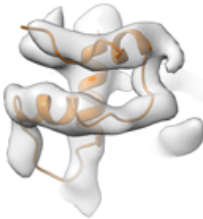
Extended Data Figure 10 | Flexibility of the enzymatic HAT and DUB modules

a, Two dimensional classes of SAGA side views revealing different positions of the enzymatic HAT and DUB modules. In the lower panel, the approximate positions of the modules are highlighted in yellow for the HAT module and in blue for the DUB module. The HAT module seems more diffuse and flexible than the DUB module whose position fluctuates less. **b**, Docking of the DUB crystal structure (pdb 6AQR) into the 15 Å resolution cryo-EM map reconstructed from a small class of particles that probably share a similar position for the DUB.



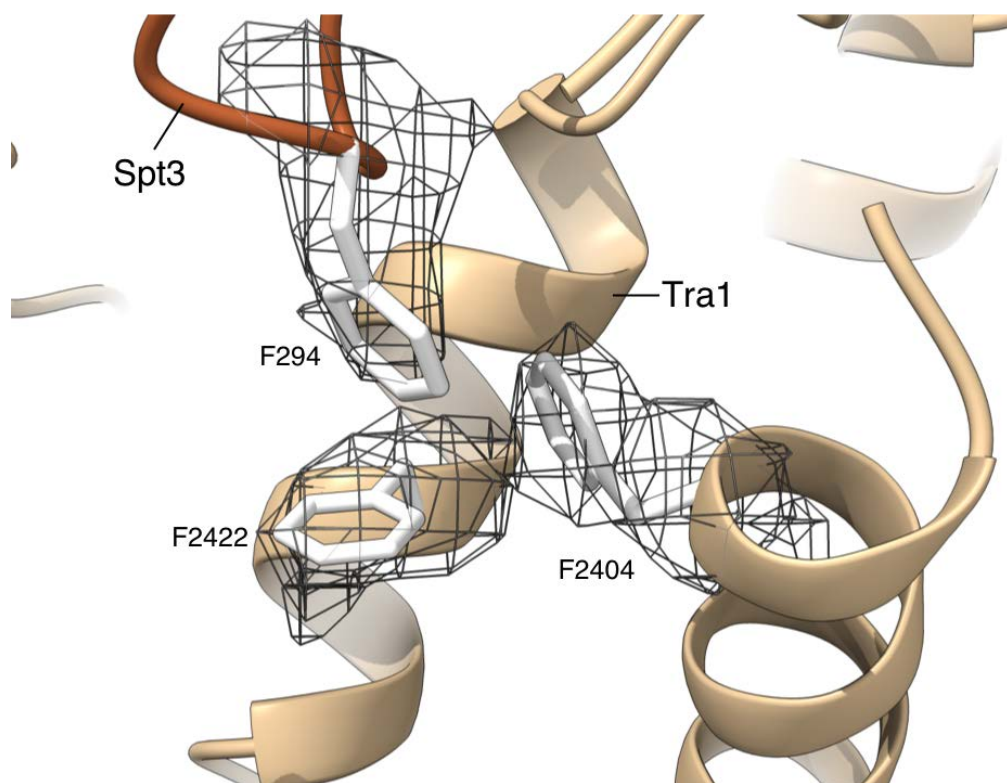
Extended Data Figure 11 | The Sgf73 linker tethers DUB to the main module.

a, Domain organization of the Sgf73 subunit. **b**, All residues that cross-link¹⁵ to the Sgf73 linker (S73L), connecting the N-terminal domain of Sgf73 (ND) present in the DUB module to the Sgf73 anchoring domain (S73A) in the central module, can be mapped on the surface of SAGA. These residues are depicted with light-blue spheres. A putative path for the 164 residues long linker is delineated (dashed) according to the cross-linking sites and scarce traces of the linker found in the cryo-EM map.

	Ada2	Ada3	
	Ada2 SWIRM aa 358-439	Ada3 (1) aa 388-472	Ada3 (2) aa 487-550
Effect of domain deletion ¹	Partially detached intact HAT	Fully detached intact HAT	Fully detached intact HAT
2D/3D structure ^{2,3}			
Xlinks to Taf6 HEAT domain ¹	(-)	(+)	(-)
Fit into EM density			

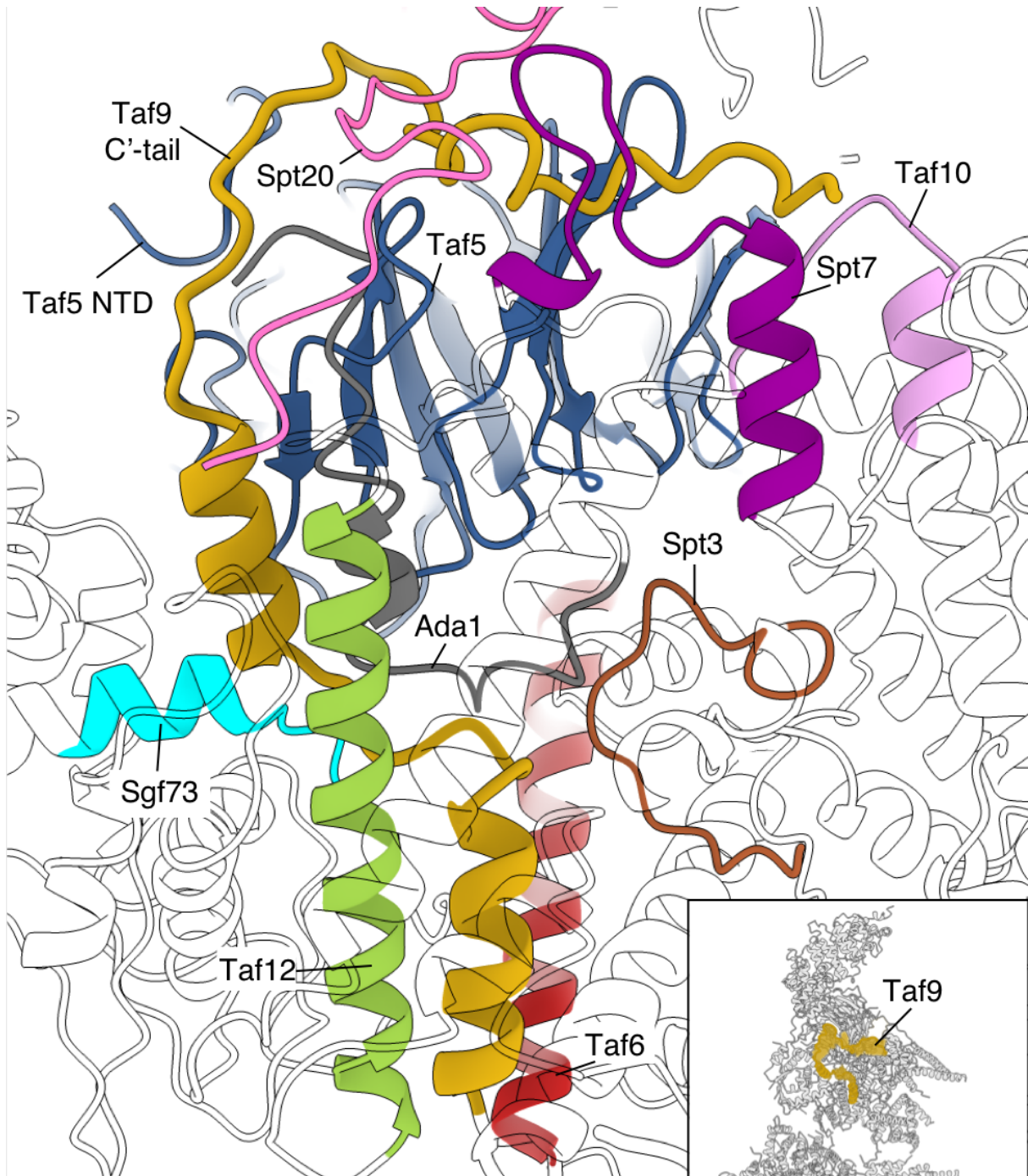
Extended Data Figure 12 | Attributing the helical domains that anchor HAT on SAGA to Ada3.

Two helical domains anchor the HAT to the Taf6 HEAT repeats in the cryo-EM maps. Domain deletion analysis identified two domains in Ada3 and one in Ada2 that serve to dock HAT on SAGA and whose deletion results in releasing intact HAT from SAGA¹⁴. Fit between the maps and secondary structure predictions or 3D structure of the Ada3 and Ada2 domains is presented in addition to available cross-linking data between these domains and the Taf6 HEAT repeats.



Extended Data Figure 13 | Interaction site of the Spt3 bridge with Tra1

The loop between helices $\alpha 2$ and $\alpha 3$ of cSpt3-HF inserts between two helices of Tra1. Aromatic stacking interactions are formed between Spt3-F294 of this loop and F2404 and F2422 of Tra1. The map density for these three phenylalanines is depicted.



Extended Data Figure 14 | Loops and tails contribute to an intricate network of interactions.

The C-terminal tail of Taf9 (R101-L139) and the short preceding loop illustrate the importance of regions lacking secondary structure elements in SAGA. These parts of Taf9 interact with many subunits and domains, including the Spt3 C'-tail, Spt20 bridge domain, Taf12-HF, Ada1, Sgf73 anchoring domain, Taf6-HF, Spt7-HF, Taf5-WD40, Taf5-NTD and Taf10-HF.

Extended Data Table 1

Cryo-EM data collection, refinement and validation statistics

	SAGA EMDB-XXXX	SAGA-TBP EMDB-XXXX	Modelled coordinates (PDB XXXX)
Data collection and processing			
Magnification	105000	105000	
Voltage (kV)	300	300	
Electron exposure (e ⁻ /Å ²)	52.8	52.8	
Defocus range (μm)	0.6-4.6	0.5-4.9	
Pixel size (Å)	1.09	1.09	
Symmetry imposed	C1	C1	
Initial particle images (no.)	426839	1068534	
Final particle images (no.)	218114	338850	
Map resolution (Å)			
Overall	5.92	3.80	
Main Lobe	4.87	3.52	
Tra1 Lobe	4.52	3.50	
FSC threshold	0.143	0.143	
Map resolution range (Å)			
Refinement			
Model composition			
Non-hydrogen atoms			39144
Protein residues			5144
Ligands			
R.m.s. deviations			
Bond lengths (Å)			0.007 (5)
Bond angles (°)			1.134 (67)
Validation			
MolProbity score			2.08
Clashscore			7.86
Poor rotamers (%)			1.43
Ramachandran plot			
Favored (%)			90.40
Allowed (%)			9.24
Disallowed (%)			0.36

