

## **Supplemental data**

### **Supplemental Figure 1**

Titration of BS3 cross-linker ratio for Pol II core complex (SDS-PAGE analysis). The individual subunits of Pol II resolved into bands as seen in lane 1 (no cross-linker) disappear with increasing concentration of cross-linker. A higher order linkage product (asterisk) most likely corresponds to a Pol II dimer (compare Figure 1b).

### **Supplemental Figure 2**

Zoom into the region of Pol II (1WCM) with inconsistency between crystallographic and cross-link data. Alpha-carbons of linked residues are highlighted by coloured sphere (hotpink for Rpb1, red for Rpb2 and firebrick for Rpb11), B-factor for linked residues are displayed under the residue label. The annotated high resolution spectra (on the right side) of cross-linked peptides corresponding to the linkage provide the experimental evidence for the cross-link data. Confidence of match has been assigned for each spectrum (Methods).

- (A) Linkage Rpb1 K343 – Rpb2 K886 observed in GNLMGK(xl)R / MK(xl)HGTYDKLDDGLIAPGVR (m/z 623.5219 5+)
- (B) Linkage Rpb1 K419 – Rpb2 K914 observed in VSGEDVIIGK(xl)TTPISPDEEELGQR / K(xl)SYR (m/z 815.6716 4+)
- (C) Linkage Rpb2 K228 – Rpb2 K246 observed in K(xl)AAPAPISHVAEIR / SALEK(xl)GSR (m/z 615.8439 4+)
- (D) Linkage Rpb1 K637 – Rpb11 K20 observed in LIKDPDTK(xl)APNAVVTIFEK / EDGPQVCAK(xl)LFGNIQK (m/z 676.3741 6+)
- (E) Linkage Rpb2 K934 – Rpb2 K972 observed in TTK(xl)IPQIGDK / TAYHSK(xl)R (m/z 525.7901 4+)
- (F) Linkage Rpb1 K343 – Rpb2 K864 observed in SYMDQEK(xl)K / GNLMGK(xl)R (m/z 647.6555 3+)
- (G) Linkage Rpb1 K941 - Rpb1 K461 observed in VK(xl)VIPYSTFR / K(xl)FLR (m/z 478.2869 4+)

### **Supplemental Figure 3**

Sequence alignment and domain architecture for Tfg1 and Tfg2.

### **Supplementary Figure 4:**

Residues on Pol II surface cross-linked to TFIIF

High-confidence cross-linking sites on Pol II surface coloured according to cross-linked TFIIF domains. The location of other TFIIF regions is indicated with circles. Two views are used, the top view and the side view, related by a 90 degree rotation around the horizontal axis. For linkage sites that are absent from the Pol II structure or the model of the Tfg1-Tfg2 dimerization domain the nearest residue that is present is highlighted and indicated with asterisk in the label. (Compare to Suppl. Table 3).

### Supplementary Figure 5:

Proximity of TFIIF dimerization domain to the Pol II lobe.

- (A) Residues for which experimental data indicate proximity to the TFIIF dimerization domain are coloured on the Pol II surface: our cross-linking analysis (blue), FeBABE by Chen *et al.*, 2007 (green) and both analyses (red). Three views are used, the top view, the front view and the side view, related by a 90 degree rotation around the horizontal and the vertical axis, respectively. The inset zooms into the region of dimerization domain binding with the labelled sites involved in cross-linking or FeBABE between the dimerization domain and Pol II. For linkage sites that are absent from the Pol II structure the nearest residue that is present is highlighted and labelled.
- (B) List of linkages (connections) indicates location of TFIIF dimerization domain on Pol II surface close to the Lobe. Data from our cross-linking analysis and Chen *et al.* (2007) are listed.

### Supplementary Figure 6:

Alternative location of TFIIF dimerization domain on top of cleft.

- (A) The dimerization domain has been placed on top of the cleft as an alternative location to the Pol II lobe as shown in figure 4 and supplementary figure 4. Two views are used, the top view and the side view, related by a 90 degree rotation around the horizontal axis. The position of DNA during transcription has been indicated.
- (B) Data in support of the alternative location. The inset zooms into the region of dimerization domain binding to Pol II with the labelled sites involved in cross-linking (red) or FeBABE (green) between the dimerization domain and Pol II. Experimentally linked residue pairs are connected by a dashed line, same colour code as for residues. Sphere highlight alpha-carbons of linked residues in the dimerization domain.
- (C) Table of experimentally linked residue pairs as displayed in panel B.

### Supplementary Figure 7:

Alternative positions in the TFIIF-Pol II complex of the Tfg2 C-terminal tail composed of linker, winged-helix (WH) domain, and C-terminal region on the Pol II surface.

- (A) Domain structure of Tfg2.
- (B) Dynamic binding patch of Tfg2 C-terminal part on the wall and clamp domains of Pol II. Linkage sites on Pol II colour coded by Tfg2 regions cross-linked to: linker (red), WH domain (purple), both (blue), and C-terminal region (green). The top view is used. Left panels show only linkages to linker and WH domain. Right panels show linkages to the C-terminal region in addition and taking priority in colouring. Small view shows important Pol II domains highlighted using the canonical colours for reference (see Figure 4C for details). The Tfg2 linker is found to cross-link over the entire protrusion and extending towards the wall and clamp domains of Pol II. This is consistent with a dynamic interaction between linker and protrusion or, less likely, multiple specific binding positions of the linker on the protrusion. The WH domain shares cross-link sites with the linker on the wall and clamp domains (Rpb2 865, Rpb1 49). All linkage sites to the linker and WH domain are shared with the C-terminal region of Tfg2 (right panel). This is consistent with a dynamic interaction of all three Tfg2 regions in the wall and clamp region of Pol II. This

binding patch is not likely to play a role in pre-initiation complexes as it collides with the position of the DNA and other transcription factors.

- (C) Specific binding position of WH domain on the protrusion on the upstream face of Pol II. Same colour coding as panel B. The side view is used, related by a 90 degree rotation around the horizontal axis to the top view used in panel B. Small view as described in panel B. None of Tfg2 domains share linkage sites: Tfg2 linker positions on external 1 domain, WH domain links to Rpb10 59 and Rpb2 813, C-terminal region links to Rpb3 137. This is consistent with a stable binding position of the Tfg2 C-terminal part in this region of Pol II, our preferred model discussed in the main body of the text.

**Supplemental Table 1**

Linked residue pairs in the Pol II core complex.

**Supplemental Table 2**

Cross-linked peptides in the Pol II core complex.

**Supplemental Table 3**

Linked residue pairs in the Pol II -TFIIF complex.

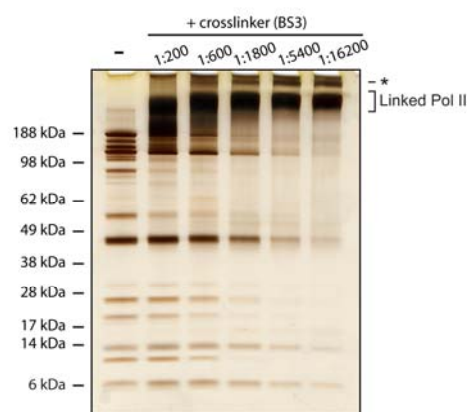
**Supplemental Table 4**

Cross-linked peptides in the Pol II-TFIIF complex.

**Supplemental Table 5**

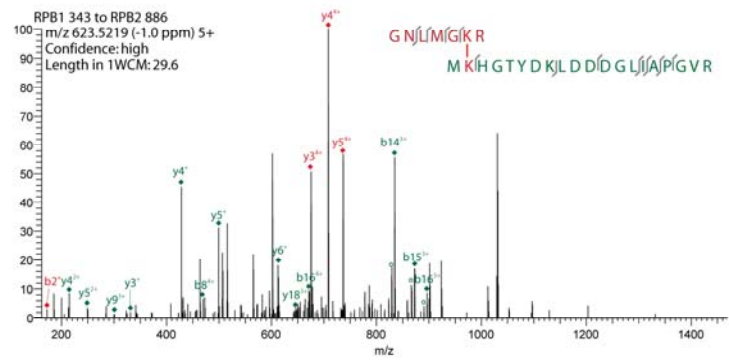
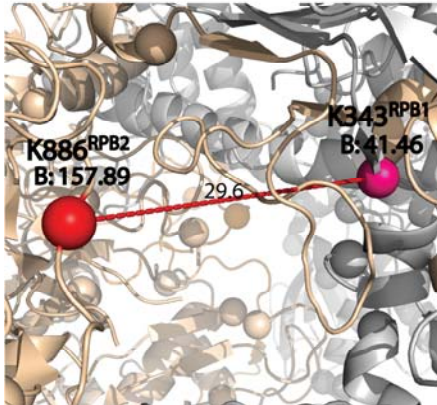
List of linked residue pairs used for positioning the TFIIF dimerization domain.

# Supplementary Figure 1



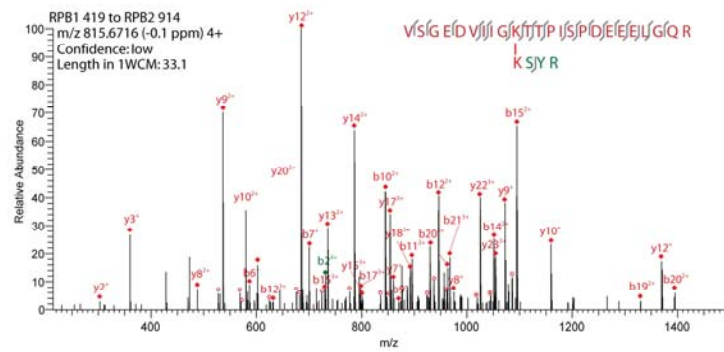
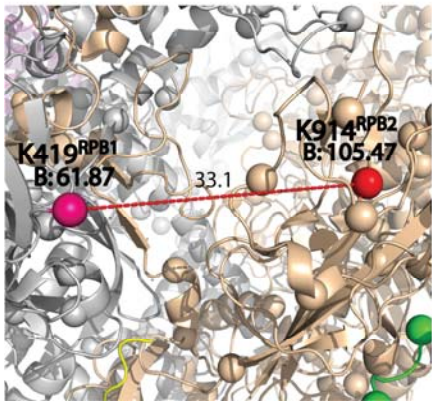
Supplementary Figure 2

A



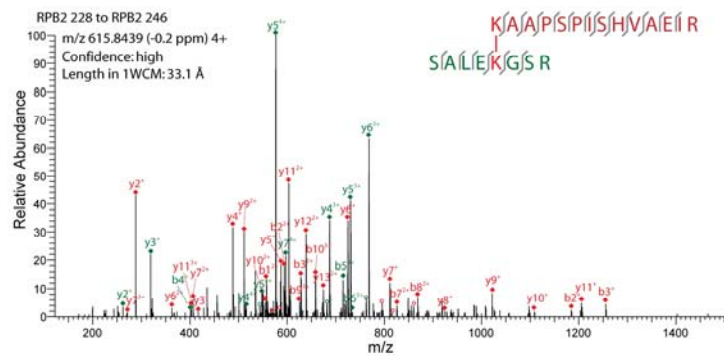
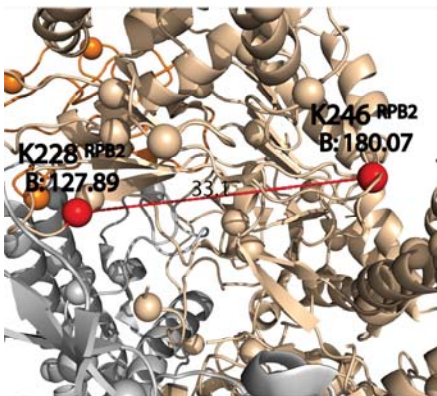
Supplementary Figure 2

B



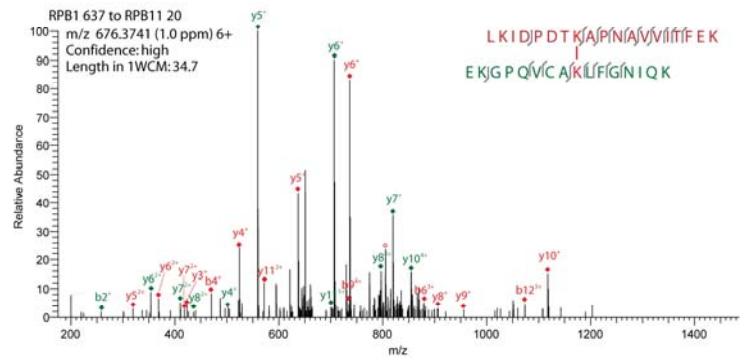
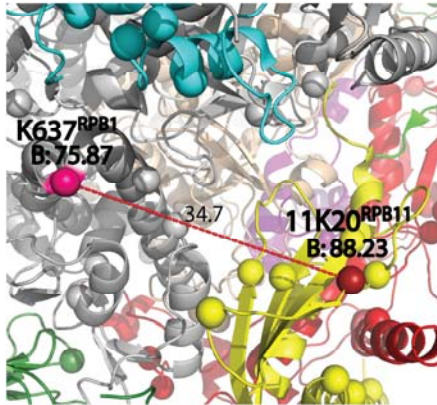
Supplementary Figure 2

C



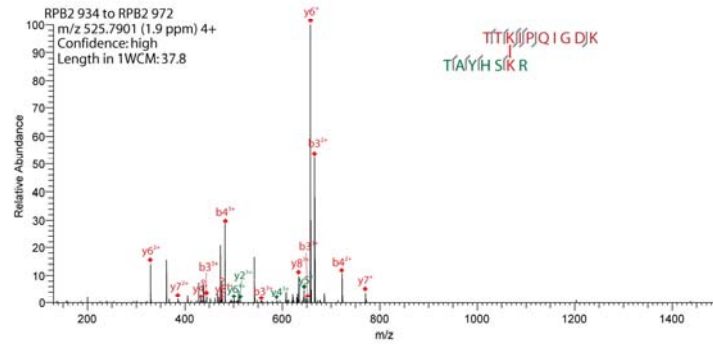
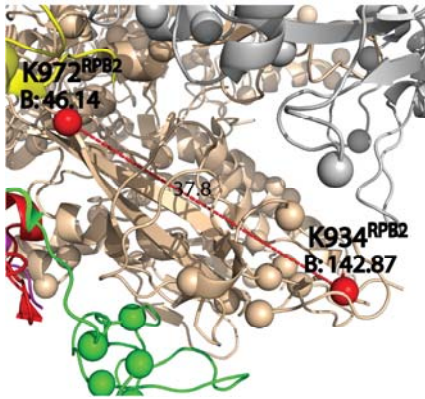
Supplementary Figure 2

D



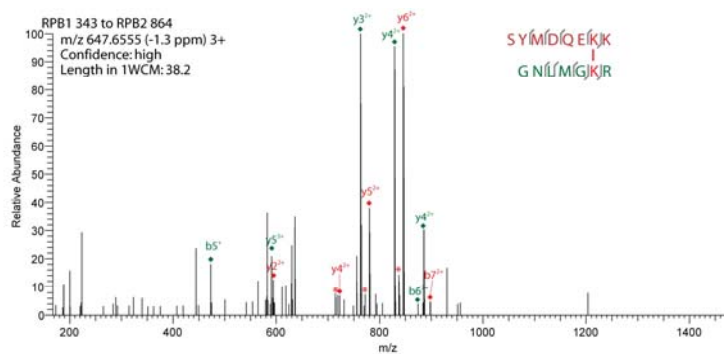
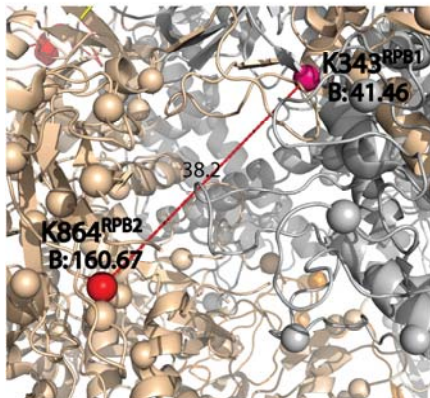
Supplementary Figure 2

E



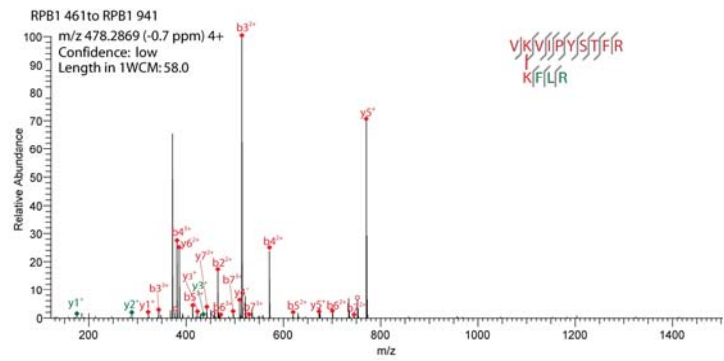
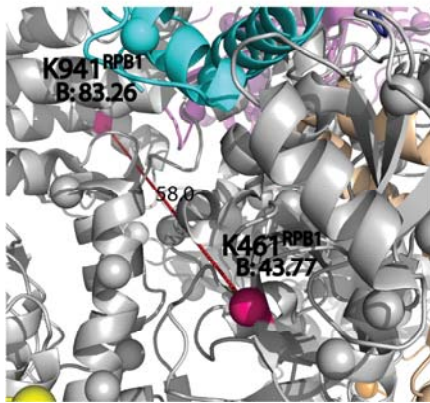
Supplementary Figure 2

F



Supplementary Figure 2

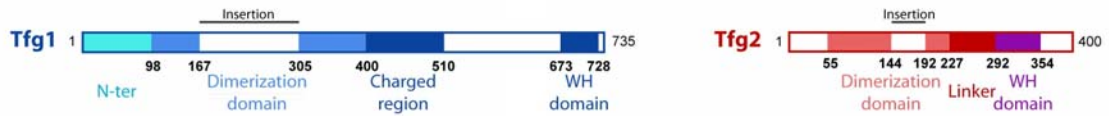
G





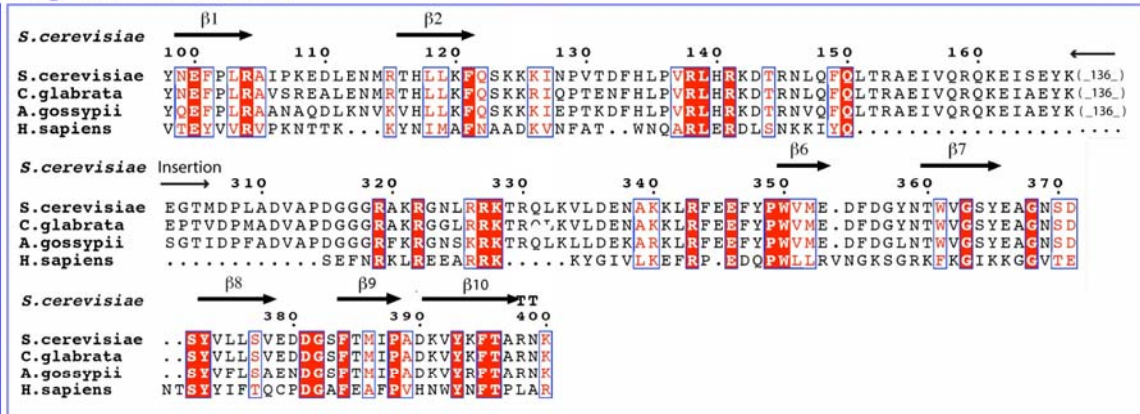
# Supplementary Figure 3

A



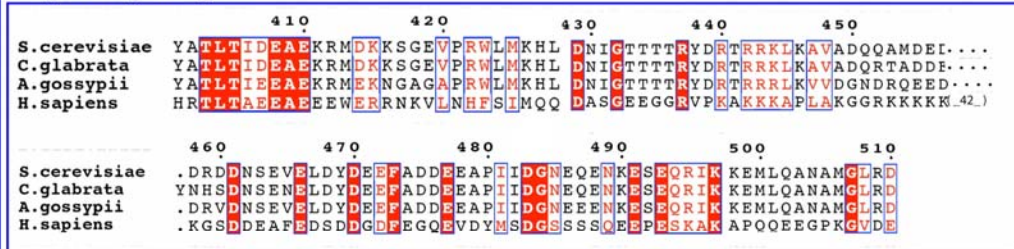
B

## Tfg1 dimerization domain

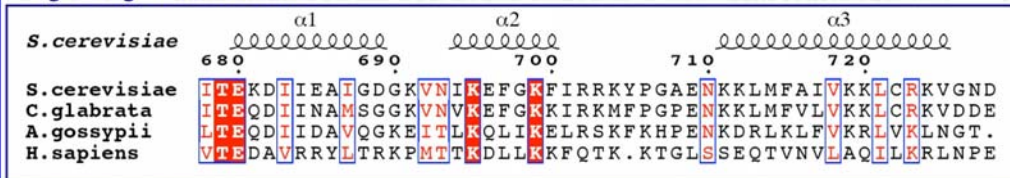


Tfg1

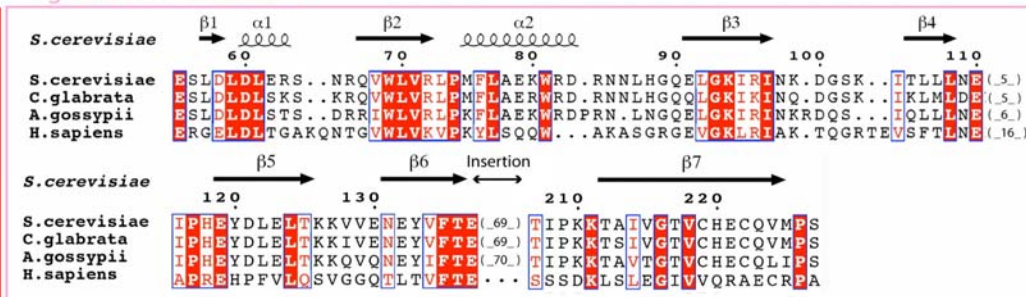
## Tfg1 charged region



## Tfg1 winged helix domain

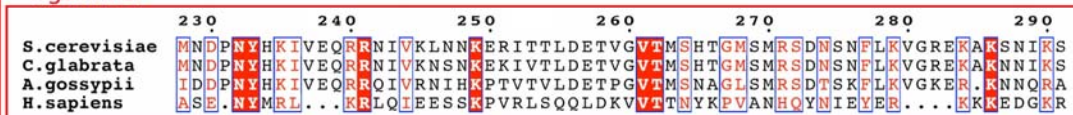


## Tfg2 dimerization domain

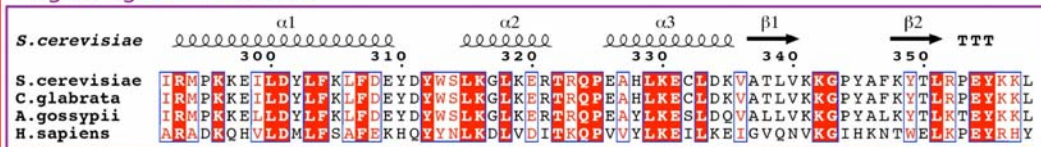


Tfg2

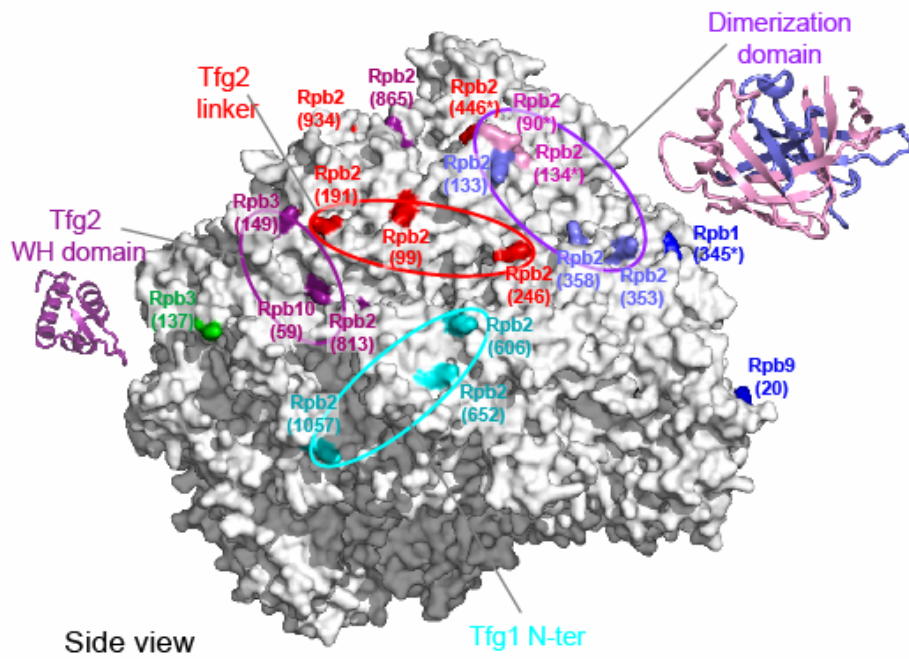
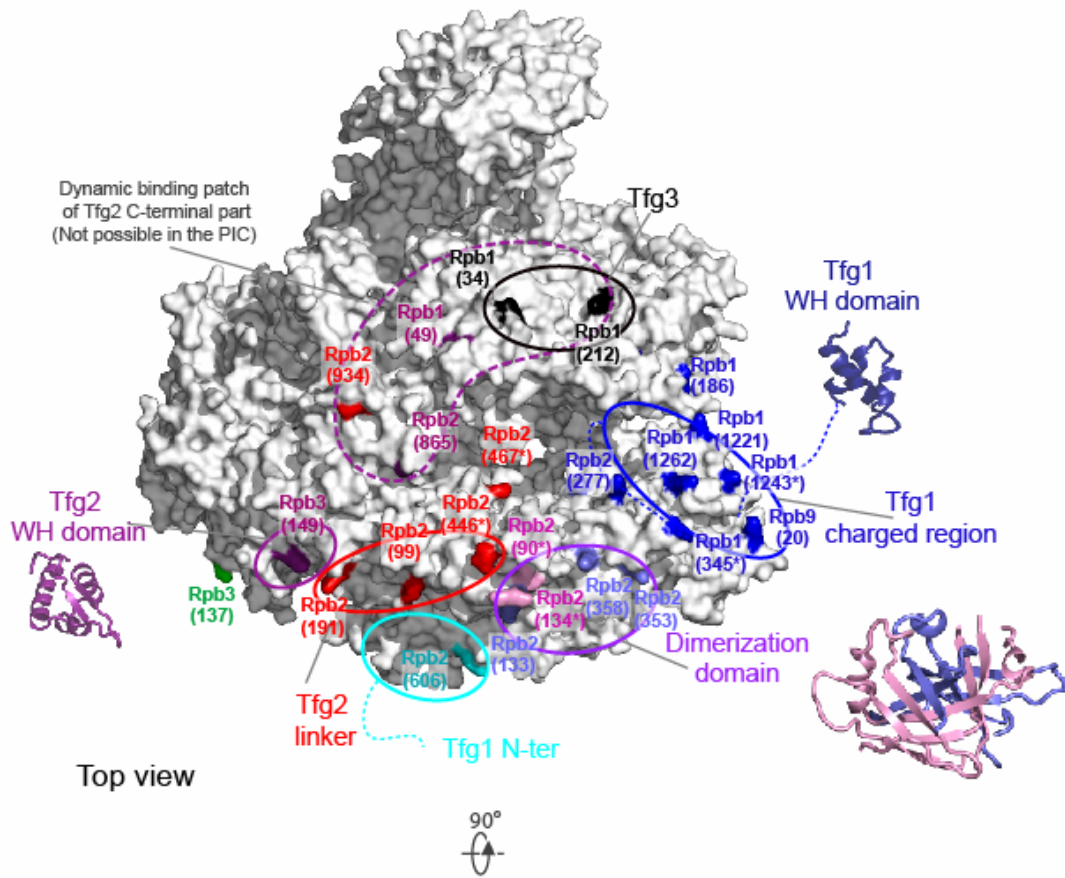
## Tfg2 linker



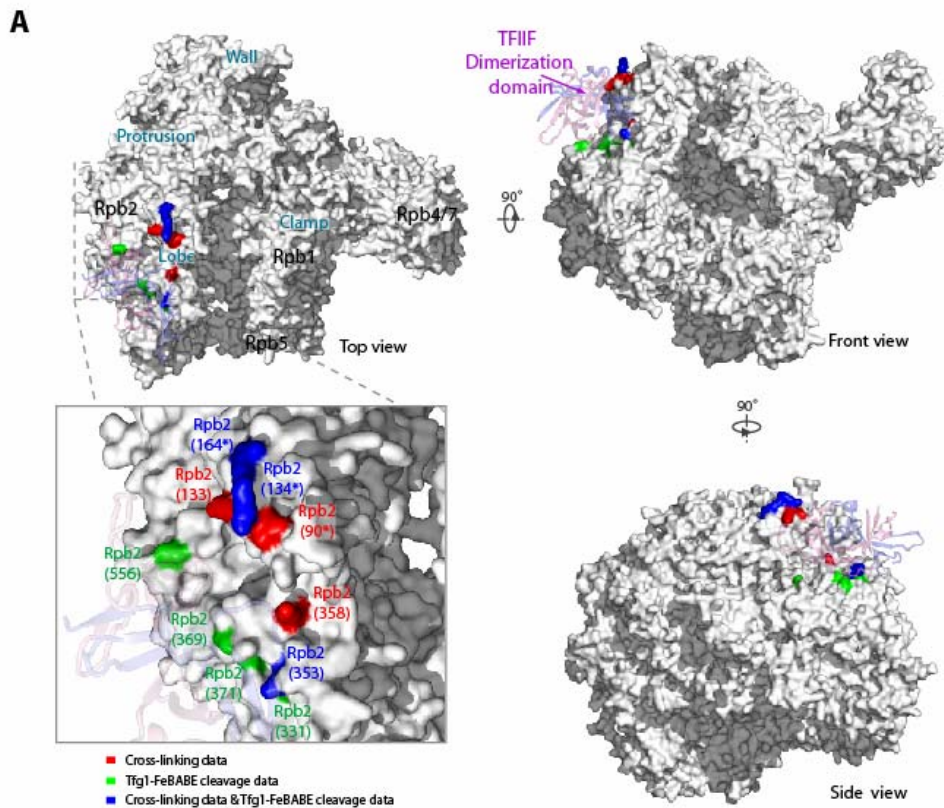
## Tfg2 winged helix domain



Supplementary Figure 4



Supplementary Figure 5



**B**

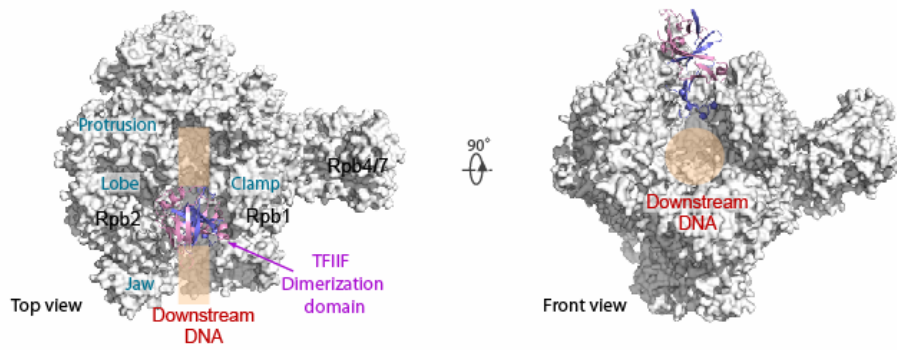
Data source <sup>[1]</sup>	Protein 1	Residue 1	Protein 2	Residue 2	Nearest residue in PDB1WCM
Cross-linking	Tfg1	Lys394	Rpb2	Lys353	
Cross-linking	Tfg1	Lys340	Rpb2	Lys87	Ile90
Cross-linking	Tfg1	Lys340	Rpb2	Lys358	
Cross-linking	Tfg2	Lys142	Rpb2	Lys87	Ile90
Cross-linking	Tfg1	Lys333	Rpb2	Lys87	Ile90
Cross-linking	Tfg1	Lys333	Rpb2	Lys133	
Cross-linking	Tfg1	Lys333	Rpb2	Lys358	
Cross-linking	Tfg2	Lys142	Rpb2	Lys148	Lys134, Lys164
Tfg1-FeBABA cleavage	Tfg1	Thr329(Thr 334) <sup>[2]</sup>	Rpb2	Leu331	
Tfg1-FeBABA cleavage	Tfg1	Thr329(Thr 334) <sup>[2]</sup>	Rpb2	Lys353	
Tfg1-FeBABA cleavage	Tfg1	Thr329(Thr 334) <sup>[2]</sup>	Rpb2	Glu371	
Tfg1-FeBABA cleavage	Tfg1	Asn364(Asn 369) <sup>[2]</sup>	Rpb2	Ile152	Lys134, Lys164
Tfg1-FeBABA cleavage	Tfg1	Asn364(Asn 369) <sup>[2]</sup>	Rpb2	Gly369	
Tfg1-FeBABA cleavage	Tfg1	Asn364(Asn 369) <sup>[2]</sup>	Rpb2	Thr556	

<sup>[1]</sup> FeBABA cleavage data from Chen *et al.* (2007) *Nat Struct Mol Biol* 8: 696-703.

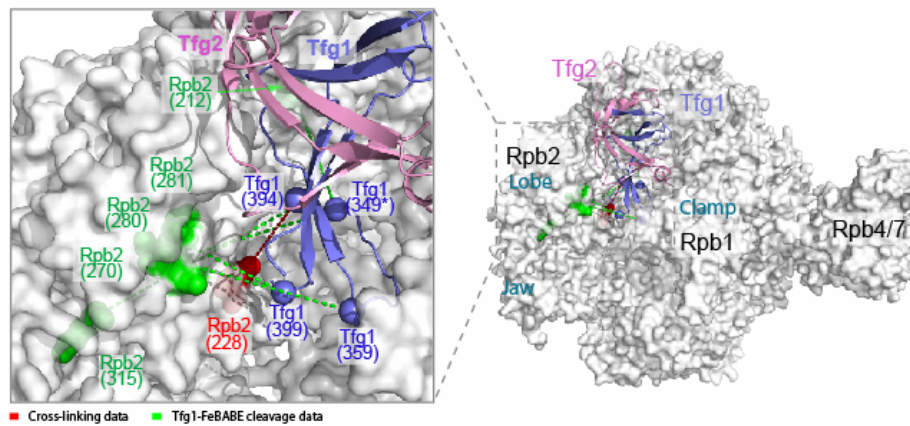
<sup>[2]</sup> Residue in *S.mik* sequence. The corresponding residue in *S.cer* used for our homology model shown in parenthesis. Alignment according to Chen *et al.* (2007) *Nat Struct Mol Biol* 8: 696-703.

Supplementary Figure 6

**A**



**B**



**C**

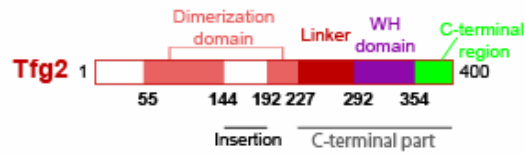
Data source <sup>[1]</sup>	Protein 1	Residue 1	Nearest residue in TFIIF dimerization domain homology model	Protein 2	Residue 2
Cross-linking	Tfg1	Lys394		Rpb2	Lys228
Tfg1-FeBABE cleavage	Tfg1	Val329(Val334) <sup>[2]</sup>	Pro349	Rpb2	Lys270
Tfg1-FeBABE cleavage	Tfg1	Tyr343(Tyr348) <sup>[2]</sup>	Pro349	Rpb2	Leu212
Tfg1-FeBABE cleavage	Tfg1	Asn354(Asn359) <sup>[2]</sup>		Rpb2	Pro281
Tfg1-FeBABE cleavage	Tfg1	Ser359(Ser364) <sup>[2]</sup>		Rpb2	Lys315
Tfg1-FeBABE cleavage	Tfg1	Asn394(Asn399) <sup>[2]</sup>		Rpb2	Ile280

<sup>[1]</sup> FeBABE cleavage data from Chen *et al.* (2007) *Nat Struct Mol Biol* 8: 696-703.

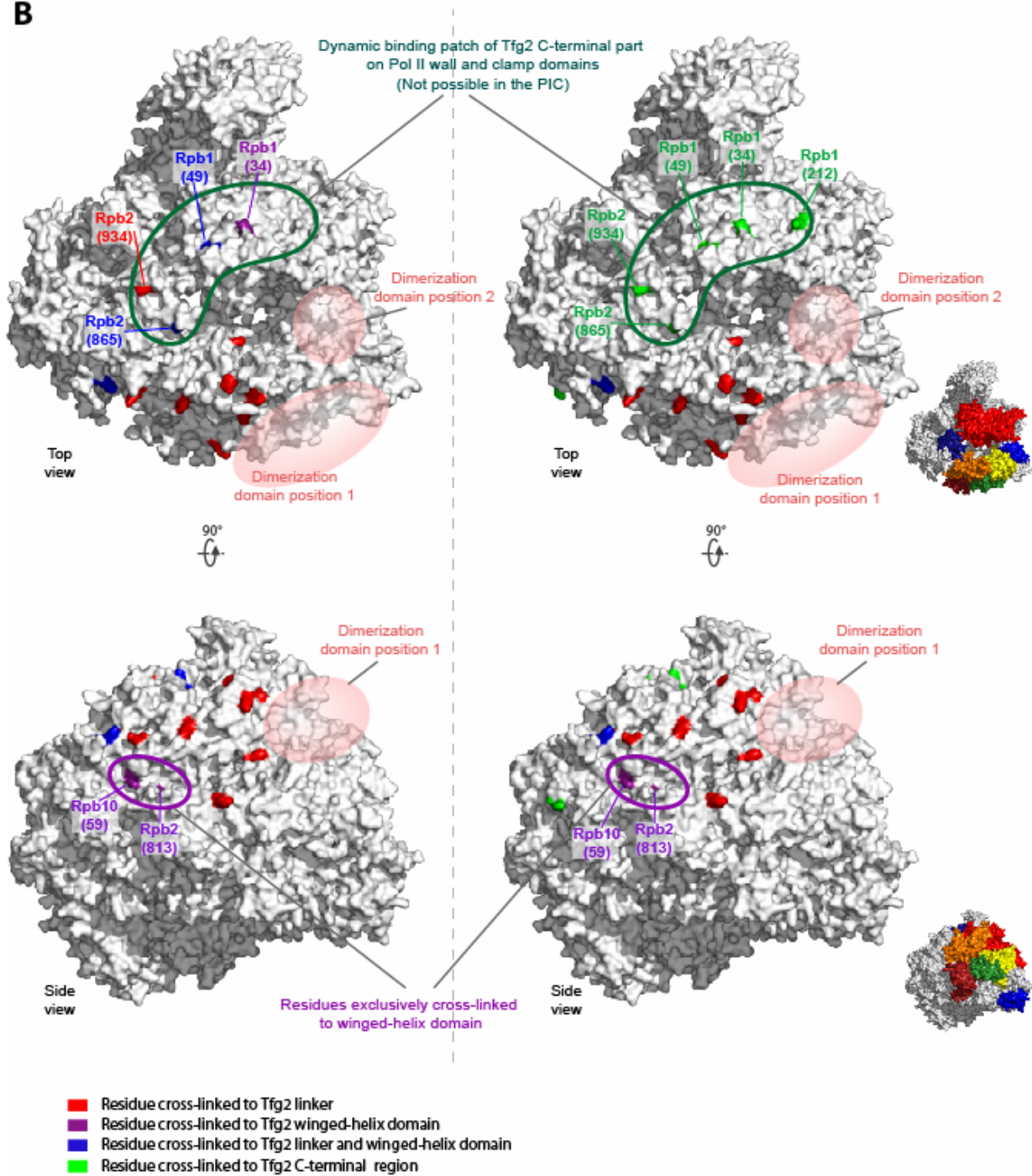
<sup>[2]</sup> Residue in *S.mik* sequence. The corresponding residue in *S.cer* used for our homology model shown in parenthesis. Alignment according to Chen *et al.* (2007) *Nat Struct Mol Biol* 8: 696-703.

# Supplementary Figure 7

**A**



**B**



## Supplementary Table 5

Protein 1	Residue 1	Nearest residue in TFIIIF dimerization domain homology model	Protein 2	Residue 2	Nearest residue in PDB 1WCM
Tfg1	Lys394		Rpb2	Lys353	
Tfg1	Lys340	Pro349	Rpb2	Lys87	Ile90
Tfg1	Lys340	Pro349	Rpb2	Lys358	
Tfg2	Lys142	Gln138	Rpb2	Lys87	Ile90
Tfg2	Lys206	Lys211	Rpb2	Lys344	Lys345
Tfg1	Lys333 <sup>[1]</sup>	Pro349	Rpb2	Lys87	Ile90
Tfg1	Lys333 <sup>[1]</sup>	Pro349	Rpb2	Lys133	
Tfg1	Lys333 <sup>[1]</sup>	Pro349	Rpb2	Lys358	
Tfg1	Lys333 <sup>[1]</sup>	Pro349	Rpb2	Lys422	
Tfg1	Lys333 <sup>[1]</sup>	Pro349	Rpb2	Lys426	
Tfg2	Lys142	Gln138	Rpb2	Lys148 <sup>[1]</sup>	Lys134, Lys164

[1]. This residue is absent in the Pol II structure or the model of Tfg1-Tfg2 dimerization domain. The nearest residue that is present is more than 10 residues away in sequence. The linkage involving this residue is not used for positioning, but is used for validating the location.