

ARTICLE

Received 16 Mar 2014 | Accepted 5 Jun 2014 | Published 10 Jul 2014

DOI: 10.1038/ncomms5310

# Conserved architecture of the core RNA polymerase II initiation complex

Wolfgang Mühlbacher<sup>1,\*,†</sup>, Sarah Sainsbury<sup>1,\*,†</sup>, Matthias Hemann<sup>1</sup>, Merle Hantsche<sup>1</sup>, Simon Neyer<sup>1,†</sup>, Franz Herzog<sup>1</sup> & Patrick Cramer<sup>1,†</sup>

During transcription initiation at promoters of protein-coding genes, RNA polymerase (Pol) II assembles with TBP, TFIIB and TFIIF into a conserved core initiation complex that recruits additional factors. The core complex stabilizes open DNA and initiates RNA synthesis, and it is conserved in the Pol I and Pol III transcription systems. Here, we derive the domain architecture of the yeast core pol II initiation complex during transcription initiation. The yeast complex resembles the human initiation complex and reveals that the TFIIF Tfg2 winged helix domain swings over promoter DNA. An ‘arm’ and a ‘charged helix’ in TFIIF function in transcription start site selection and initial RNA synthesis, respectively, and apparently extend into the active centre cleft. Our model provides the basis for further structure–function analysis of the entire transcription initiation complex.

<sup>1</sup>Gene Center Munich and Department of Biochemistry, Center for Integrated Protein Science CIPSM, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. \* These authors contributed equally to this work. † Present address: Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany. Correspondence and requests for materials should be addressed to F.H. (email: herzog@LMB.uni-muenchen.de) or to P.C. (email: cramer@LMB.uni-muenchen.de).

**D**uring assembly of the transcription preinitiation complex (PIC), the RNA polymerase (Pol) II–TFIIF complex binds to a TFIIB–TBP–DNA promoter assembly, resulting in a core initiation complex<sup>1,2</sup>. The structure and function of the core initiation complex is conserved from yeast to human, and also in the two other eukaryotic transcription systems<sup>3</sup>. Pol I and Pol III both contain a TFIIF-like subcomplex, and they also use TBP and a TFIIB-like factor for initiation. The conserved core initiation complex stabilizes open promoter DNA and directs initial RNA synthesis, resulting in the initially transcribing complex (ITC).

In the Pol II system, the core initiation complex additionally binds TFIIE and TFIIH to form a complete PIC. Architectural models of the yeast Pol II PIC were obtained by site-specific protein cleavage mapping<sup>4–6</sup>. The architecture of the human PIC was obtained by electron microscopy (EM)<sup>7</sup>, and generally resembled that of the yeast PIC. Recently, an alternative model of the yeast PIC was derived based on a combination of EM and protein crosslinking coupled to mass spectrometry (XL-MS)<sup>8</sup>, raising the question whether the PIC architecture is indeed conserved between eukaryotic species.

We have previously modelled the architecture of the core Pol II initiation complex<sup>9</sup> by structural superposition of our Pol II–TFIIB crystal structures<sup>10,11</sup> with a Pol II–TFIIF complex model obtained by XL-MS<sup>12</sup>. However, the model awaited experimental confirmation because both TFIIF and TFIIB are modular factors with flexible domains that may be repositioned on complex assembly. The amino-terminal (N-terminal) regions of TFIIF subunits Tfg1 and Tfg2 form a dimerization module, whereas their flexibly linked carboxy-terminal (C-terminal) regions each include a winged helix (WH) domain. TFIIB consists of a N-terminal zinc ribbon domain followed by the reader and linker regions and two C-terminal cyclin domains.

Here, we used protein–protein crosslinking and mass spectrometric identification of the crosslinked sites to derive a model of the core ITC from yeast. The yeast complex resembles the previously published human counterpart<sup>7</sup>, indicating that the core initiation complex is conserved between eukaryotic species. The results also reveal a new element, the charged helix in the TFIIF subunit Tfg1, demonstrate that the Tfg2 WH domain can swing over the DNA after it was loaded into the active centre cleft, and provide a basis for elucidating the architecture of the entire initiation complex.

## Results

**Preparation and crosslinking analysis of the core ITC.** To experimentally define the domain architecture of the core initiation complex, we reconstituted a defined yeast core ITC. We previously noted that a stable, defined ITC can be formed by including a 6 nt RNA product<sup>11</sup>. We therefore incubated purified Pol II, TFIIF, TFIIB and TBP with a DNA–RNA scaffold (Fig. 1a) and obtained a stable core ITC after size exclusion chromatography (Fig. 1b; see Methods). We then analysed this complex by XL-MS. The power and reliability of XL-MS was recently exemplified by a high agreement between Pol I models derived from XL-MS<sup>13</sup> and subsequently from X-ray analysis<sup>14</sup>.

We obtained a total of 472 high-confidence lysine–lysine protein crosslinks (332 distance restraints) within the core ITC, of which 241 were intersubunit and 231 were intrasubunit crosslinks (Table 1). A total of 194 crosslinks within Pol II were readily explained with the Pol II crystal structure<sup>15</sup>. Another 33 crosslinks were observed between TFIIF subunits Tfg1 and Tfg2, and could be explained with the TFIIF dimerization module structure<sup>16</sup>. Only 18 crosslinks showed  $C\alpha$  distances above the maximum expected distance of  $27 \pm 3 \text{ \AA}$ <sup>12</sup> (Fig. 1c). Most of these could be explained by the known structural flexibility and higher

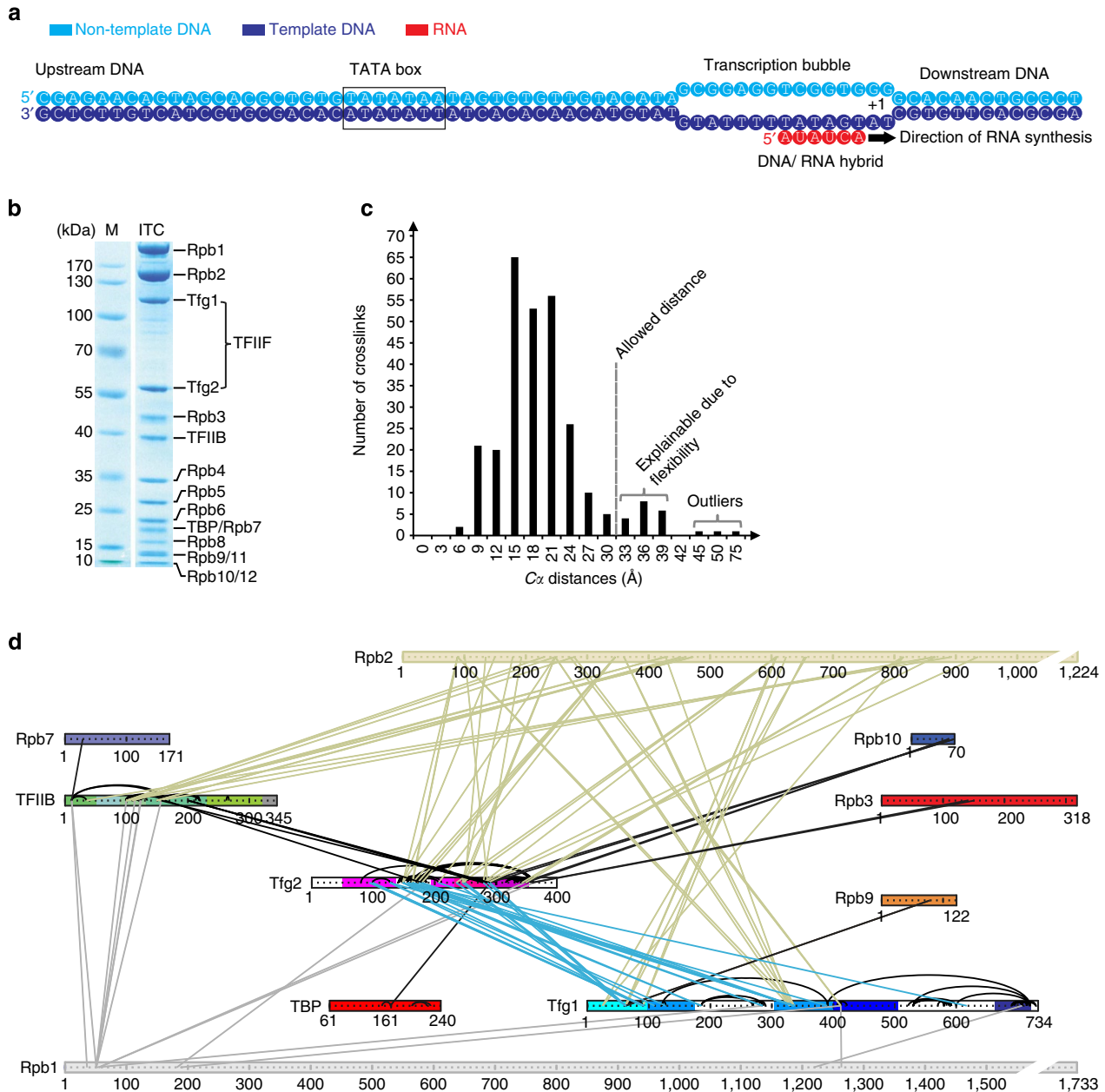
crystallographic B-factors of the involved lysine residues, leaving only three crosslinks unexplained. Within TFIIB and TBP, 23 and six intramolecular crosslinks were observed, respectively, and could be explained with crystal structures<sup>10,11,17,18</sup>. These multiple internal controls demonstrate the high reliability of the observed crosslinking data.

**Positions of TFIIB and TFIIF.** We also observed 111 intermolecular crosslinks between transcription factors and Pol II (Table 1) that allowed us to model the core yeast ITC (Fig. 2a). Of these crosslinks, 21 were observed between the TFIIF dimerization module and the Pol II lobe and protrusion domains, showing that the module remains at its location observed in the binary Pol II–TFIIF complex<sup>5,12</sup>. Further, 16 crosslinks were obtained between the TFIIB linker and the Pol II domains clamp core, protrusion, and wall. Another six crosslinks were detected between the N-terminal TFIIB cyclin domain and the clamp core, protrusion and wall. The C-terminal cyclin domain did not crosslink to Pol II, consistent with its mobility<sup>10</sup>. All TFIIB–Pol II crosslinks were explained with our crystal structures of the Pol II–TFIIB complex<sup>10,11</sup>, demonstrating that TFIIB binds Pol II as observed in the binary structure. These results were corroborated by crosslinks between the TFIIF subunit Tfg2 to TBP (one crosslink) and to the N-terminal cyclin domain of TFIIB (six crosslinks).

**The Tfg2 WH domain swings over DNA in the cleft.** The crosslinking data also revealed that in the reconstituted core ITC the WH domain in Tfg2 can reside at a position near upstream DNA on the outside of Pol II<sup>12</sup>, but also at a position above the DNA in the active centre cleft (Fig. 2a,d). Thus, in the core ITC, the WH domain remains flexible and adopts both alternative positions. The original WH position near upstream DNA<sup>12</sup> gives rise to 13 crosslinks to the Pol II subunits Rpb2, Rpb3 and Rpb10. The new position above the Pol II cleft was defined by four crosslinks of the WH domain to the clamp, TBP, and the TFIIB N-terminal cyclin domain. These restraints can be satisfied when one assumes a position of the WH domain with respect to promoter DNA that resembles that in a known X-ray structure of a WH domain bound to DNA<sup>19</sup>. This position is also consistent with a recent mapping of the DNA-binding face of the Tfg2 WH domain<sup>20</sup>.

These results indicate that the Tfg2 WH domain can swing over promoter DNA after its loading into the Pol II cleft, and indicate a role of this domain in DNA melting and/or stabilization of the open complex and the ITC. Indeed, this domain binds DNA and is required for initiation<sup>21</sup>, and TFIIF suppresses abortive initial transcription<sup>22</sup>. The proximity of the Tfg2 WH domain to TFIIB indicates how TFIIF could stabilize TFIIB on Pol II during initial transcription<sup>23</sup>. The position of the Tfg2 WH domain above the cleft apparently represents its position in a complete ITC. This position is near TFIIE and TFIIA in the PIC<sup>7,8</sup>, and is likely stabilized on TFIIE and/or TFIIA binding. The other WH domain in TFIIF subunit Tfg1 only gave rise to a single crosslink at the Pol II jaw, and does not adopt a defined location<sup>12</sup>.

**Model of the yeast core ITC.** On the basis of the large number of protein crosslinks, we built a reliable three-dimensional model of the yeast Pol II core ITC. First, we derived a homology model of the yeast TFIIF dimerization module based on the human crystal structure<sup>16</sup>. Second, we positioned the resulting yeast TFIIF dimerization module model onto the Pol II–TFIIB–DNA–RNA crystal structure<sup>11</sup> assuming the location of the human



**Figure 1 | Preparation and XL-MS analysis of the yeast core ITC.** (a) DNA-RNA scaffold based on a *HIS4* DNA promoter with a mismatched bubble region containing a 6 nt RNA transcript forming a hybrid duplex with the DNA template strand<sup>11</sup>. (b) SDS-polyacrylamide analysis of the purified Pol II ITC revealing its 16 polypeptide subunits. (c)  $C\alpha$  distance distribution for observed lysine-lysine crosslink pairs (unique distance restraints). Crosslinks with distances of 30–39 Å are explainable due to protein mobility (four crosslinks) or because of lysine location in mobile protein loops with high crystallographic B-factors (14 crosslinks). Only three crosslinks cannot be explained and are classified as outliers. (d) Crosslink map of the ITC. Crosslinks within Pol II were excluded for clarity. TFIIB and Pol II subunits are colour coded as before<sup>11</sup> and TBP and TFIIF were coloured as in Fig. 2. The map was generated with a MATLAB script (Supplementary Files).

module detected by EM<sup>7</sup>. Third, we extended DNA both upstream and downstream using standard B-form duplexes.

**TFIIF arm and charged helix.** In the resulting model, the Tfg1 ‘arm’ (a  $\beta$ -hairpin comprising yeast residues 146–153 and 319–338) extends from the TFIIF dimerization module, traversing between the Pol II protrusion and lobe domains into the active centre cleft (Fig. 2c). The arm forms 19 crosslinks in the cleft, consistent with detection of the arm in the human PIC by EM<sup>7</sup>. A mutation at the point where the arm

extends from the dimerization module leads to shifts in the transcription start site<sup>24</sup>.

A second extension from the TFIIF dimerization module, a negatively charged, flexible<sup>16</sup>  $\alpha$ -helix at the beginning of the ‘charged region’ in Tfg1 (named here the ‘charged helix’, yeast residues 406–417) clashed with the Pol II lobe. A reorientation of the charged helix towards the jaws released this clash and explained six crosslinks of the charged helix to the Pol II cleft. The location of the charged helix explained a distinct EM difference density that was hypothesized to stem from the corresponding human TFIIF region<sup>7</sup>.

**Table 1 | Observed lysine-lysine crosslinks in the yeast core Pol II ITC.**

ITC parts	Crosslinks (all)	Crosslinks (mapable)
All (inter and intra crosslinks)	472	328
<i>Inter crosslinks</i>		
All	241	164
Pol II–Pol II	90	90
Pol II–TFIIF	84	40 (48)*
Pol II–Tfg1/2 dimerization module <sup>†</sup>	25 (29)	25 (29)
Pol II–Tfg1 WH	1	1
Pol II–Tfg2 WH	14 (17)	14 (17)
Pol II–TFIIB	27	23
Tfg1–Tfg2	33	11
TFIIB–Tfg2	6	0 (1)
TBP–Tfg2	1	0 (1)
<i>Intra crosslinks</i>		
All	231	164
Pol II	104	104
TFIIF	98	32
Tfg1/2 dim.-module <sup>†</sup>	4	4
Tfg1 WH model	18	18
Tfg2 WH model	11	11
TFIIB	23	22
TBP	6	6

Pol, RNA polymerase; TBP, TATA-binding protein.  
 Provided is the number of crosslinks between certain parts of the ITC, referring to unique distance restraints.  
 \*Numbers in brackets include crosslinks that involve amino acids located no more than three residues away from residues within known structures.  
 †Residues from the charged region of Tfg1<sup>12</sup> (400–417) and N-terminal region (92–98) are also part of the dimerization model based on the human X-ray structure<sup>16</sup>.

Published mutational and kinetic data revealed important roles of the charged helix in transcription initiation and elongation<sup>25,26</sup>. These roles can now be rationalized due to the location near downstream DNA (Fig. 2c). The charged helix apparently repels the downstream DNA from the lobe, positioning it along the clamp head on the opposite side of the cleft. This may help to stabilize melted DNA and to align the DNA template in the active site and account for the known role of the charged helix in stimulating initial RNA synthesis<sup>27</sup>.

## Discussion

Our crosslinking data and detailed architectural model of the core yeast ITC agree with previous site-specific protein cleavage mapping of the yeast PIC<sup>4,5,28</sup>. Our model further agrees with structural analysis of human Pol II PIC intermediates by EM<sup>7</sup>. Thus, the architecture of the core ITC is highly conserved between yeast and human. Domains in TFIIF and TFIIB adopt very similar locations on the Pol II surface in both species, although the position of the Tfg2 WH domain above the cleft may change slightly on binding TFIIA and/or TFIIE or different DNA template sequences. Furthermore, two TFIIF motifs extending from the dimerization module, the arm and charged helix, adopt the same locations in the downstream cleft. Our results also indicate that the overall domain architecture of the initiation complex is generally maintained during the transition from a PIC to an ITC.

Finally, our core ITC model can explain the XL-MS data obtained recently with a complete yeast PIC<sup>8</sup>. In the latter study, 117 distance restraints were obtained for Pol II, TFIIB, TFIIF and TBP. Of these, only one crosslink disagrees with our model, which was derived from 472 crosslinks with 332 distance restraints. Apparently the published study<sup>8</sup> contains correct crosslink information, but conflicting electron microscopic results, which have apparently led to an alternative initiation complex model. With respect to the core ITC, the discrepancies

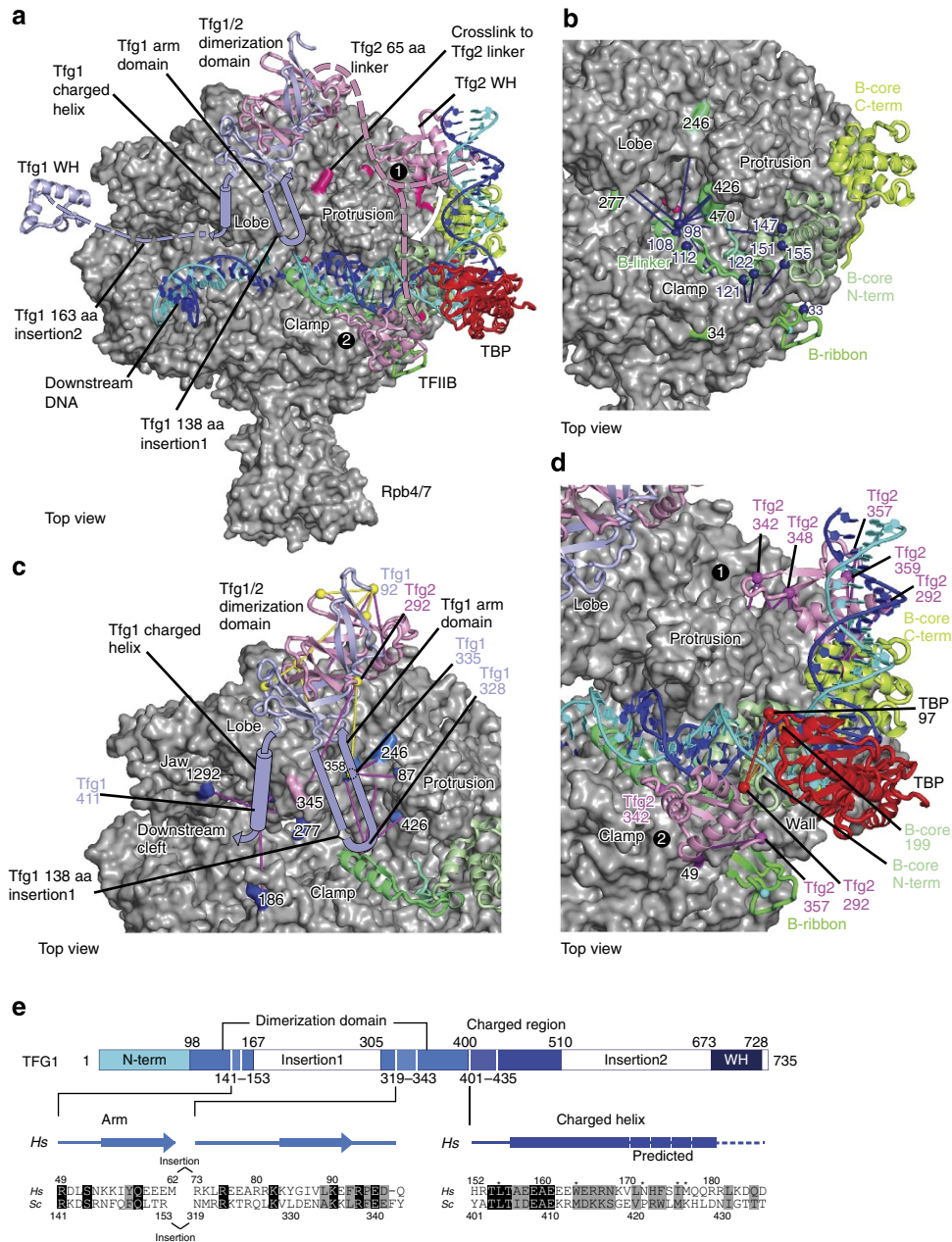
are now resolved. Our results lead to a unified, highly conserved architecture of the core transcription initiation complex. The location of the remaining general transcription factors TFIIE and TFIIF differs to some extent in three published studies<sup>6–8</sup> and may be analysed in the future.

## Methods

**Preparation of the yeast core Pol II ITC.** Endogenous *S. cerevisiae* 12-subunit Pol II was prepared as described<sup>29</sup>. Full-length TFIIB<sup>11</sup>, TFIIF (*S. mikatae* Tfg1, *S. cerevisiae* Tfg2)<sup>4</sup> and TBP<sup>30</sup> (residues 61–240) were prepared as described. Pol II (0.77 mg, 3.5 mg ml<sup>−1</sup>) was incubated with a fourfold molar excess of TFIIF, TFIIB and TBP, and a twofold molar excess of DNA–RNA scaffold (Fig. 1a) for 30 min at 298 K and for 5 min at 293 K, 288 K and 283 K. Size-exclusion chromatography in 250 mM KCl, 20 mM HEPES pH 7.5, 5% glycerol and 2 mM Dithiothreitol (DTT) resulted in a stoichiometric ITC (Fig. 1b).

**Crosslinking and mass spectrometry.** 0.9 mg purified ITC (1.2 mg ml<sup>−1</sup>) was incubated with an eightfold molar excess of DNA–RNA scaffold and crosslinked with 0.6 mM isotope-labeled disuccinimidyl suberate (DSS-d0/d12, Creative Molecules Inc.) as described<sup>13</sup>. Crosslinked protein was digested, and the crosslinked peptides were enriched, analysed by liquid chromatography coupled to tandem mass spectrometer (Orbitrap Elite), and spectra were searched by the xQuest software<sup>31,32</sup>. The resulting cross-link identifications were manually validated and the local false discovery rates for each individual cross-link were estimated as described (Supplementary Table 1). The term ‘crosslink’ describes a peptide–peptide pair linked through two specific lysines. A single peptide–peptide combination of peptides containing more than one lysine each can be identified by distinct cross-links which represent a single unique distance restraint. Finally, we also detected 11 crosslinks, which linked to serine, threonine and tyrosine with maximum C $\alpha$  distances of 21.7 Å (Supplementary Table 2). Nevertheless, these crosslinks presented no additional structural information.

**Structural modelling.** All modelling was done manually. To generate the ITC model, we used PyMOL and crosslinking restraints to place homology models of the yeast TFIIF dimerization module and WH domains and the TFIIB C-terminal cyclin domain onto the Pol II–TFIIB (PDB:4BBR) open promoter complex model<sup>10</sup>. Models for the yeast TFIIF dimerization module and WH domains and TFIIB C-terminal cyclin domain were generated from known crystal structures (dimerization module, chains A and F in PDB 1F3U; Tfg1 WH domain, PDB 1I27; Tfg2 WH domain, PDB 1BBY; TFIIB C-terminal cyclin domain, chain A in PDB 1VOL) using MODELLER<sup>33</sup>.



**Figure 2 | Crosslinking-derived model of the yeast core ITC.** (a) Top view of the ITC, highlighting the locations of TBP (red), TFIIB (green) and TFIIF subunits Tfg1 (light blue) and Tfg2 (pink) on the Pol II surface. The TFIIF arm and charged helix elements are indicated as an antiparallel  $\beta$ -hairpin and  $\alpha$ -helix, respectively. Alternative positions of the Tfg2 WH domain are indicated with black circled numbers (1, outside the cleft near upstream DNA as in the Pol II-TFIIF binary complex; 2, at the DNA bubble above the cleft). Mobile linkers are shown as dashed lines. (b) Pol II-TFIIB crosslinks (blue lines) viewed from the top as in a can be explained with the previously derived crystallographic TFIIB (b) core and ribbon domain locations<sup>10,11</sup>. (c) Location of the Pol II-TFIIF dimerization module (pink and yellow lines depict inter- and intra crosslinks, respectively). (d) The Tfg2 WH domain adopts two distinct locations. At position 1, the Tfg2 WH domain crosslinks to Pol II (pink lines), and at position 2, it crosslinks additionally to TFIIB and TBP (orange and red lines, respectively). (e) Domain organization of TFIIF subunit Tfg1 and location and conservation of the arm and charged helix elements. The charged helix was partially resolved in the X-ray structure<sup>16</sup> and is predicted to be longer<sup>34</sup>. Residues required for normal transcription initiation and elongation<sup>25</sup> are indicated as grey asterisks.

Residues 92–153 and 324–417 in *S. cerevisiae* Tfg1 correspond to residues 5–62 and 73–168 in human Rap74. Residues 54–138 and 208–227 of *S. cerevisiae* Tfg2 align to residues 2–119 of human Rap30, respectively. Residues 678–736 of *S. cerevisiae* Tfg1 WH domain align with residues 454–517 in human Rap74 and residues 292–350 in *S. cerevisiae* Tfg2 WH domain align to residues 176–243 in human Rap30. Residues 125–345 of *S. cerevisiae* TFIIB C-terminal cyclin domain align to residues 113–316 in the human counterpart. The Tfg1 sequence of *S. mikatae* was substituted with the one of *S. cerevisiae* since they only differ in three amino acids in the dimerization module and in six in its WH model.

**References**

- Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
- Grunberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* **38**, 603–611 (2013).
- Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).

4. Chen, H. T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
5. Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
6. Grunberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* **19**, 788–796 (2012).
7. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
8. Murakami, K. *et al.* Architecture of an RNA polymerase II transcription preinitiation complex. *Science* **342**, 1238724 (2013).
9. Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431–1437 (2012).
10. Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
11. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437–440 (2013).
12. Chen, Z. A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
13. Jennebach, S., Herzog, F., Aebersold, R. & Cramer, P. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* **40**, 5591–5601 (2012).
14. Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502**, 650–655 (2013).
15. Armache, K. J., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134 (2005).
16. Gaiser, F., Tan, S. & Richmond, T. J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J. Mol. Biol.* **302**, 1119–1127 (2000).
17. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
18. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl Acad. Sci. USA* **90**, 8174–8178 (1993).
19. Brent, M. M., Anand, R. & Marmorstein, R. Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure* **16**, 1407–1416 (2008).
20. Kilpatrick, A. M., Koharudin, L. M., Calero, G. A. & Gronenborn, A. M. Structural and binding studies of the C-terminal domains of yeast TFIIF subunits Tfg1 and Tfg2. *Proteins* **80**, 519–529 (2011).
21. Tan, S., Conaway, R. C. & Conaway, J. W. Dissection of transcription factor TFIIF functional domains required for initiation and elongation. *Proc. Natl Acad. Sci. USA* **92**, 6042–6046 (1995).
22. Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274**, 35668–35675 (1999).
23. Cabart, P., Ujvari, A., Pal, M. & Luse, D. S. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc. Natl Acad. Sci. USA* **108**, 15786–15791 (2011).
24. Khapersky, D. A., Ammerman, M. L., Majovski, R. C. & Ponticelli, A. S. Functions of *Saccharomyces cerevisiae* TFIIF during transcription start site utilization. *Mol. Cell Biol.* **28**, 3757–3766 (2008).
25. Funk, J. D., Nedialkov, Y. A., Xu, D. & Burton, Z. F. A key role for the alpha 1 helix of human RAP74 in the initiation and elongation of RNA chains. *J. Biol. Chem.* **277**, 46998–47003 (2002).
26. Zhang, C., Zobeck, K. L. & Burton, Z. F. Human RNA polymerase II elongation in slow motion: role of the TFIIF RAP74 alpha1 helix in nucleoside triphosphate-driven translocation. *Mol. Cell Biol.* **25**, 3583–3595 (2005).
27. Ren, D., Lei, L. & Burton, Z. F. A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly. *Mol. Cell Biol.* **19**, 7377–7387 (1999).
28. Fishburn, J. & Hahn, S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Mol. Cell Biol.* **32**, 12–25 (2012).
29. Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).
30. Treutlein, B. *et al.* Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol. Cell* **46**, 136–146 (2012).
31. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
32. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9**, 901–903 (2012).
33. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
34. Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

### Acknowledgements

We thank Clemens Plaschka (Gene Center Munich) and Ruedi Aebersold (ETH Zürich). S.N. was supported by a PhD student fellowship of the Boehringer Ingelheim Fonds. F.H. is funded by the Bavarian Research Center for Molecular Biosystems and by a LMUexcellent junior grant. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB646, SFB960, SFB1064, CIPSM, NIM, QBM, GRK1721), an ERC Advanced Grant, the Jung-Stiftung, and the Vallee Foundation.

### Author contributions

W.M. performed experiments. S.S. and W.M. analysed data. M.He and F.H. collected and validated MS data. M.Ha modelled the yeast TFIIF dimerization and winged helix modules and the yeast TFIIB C-terminal cyclin domain. S.N. programmed the MATLAB script used to generate Fig. 1d. P.C. designed research and wrote the manuscript, with help from all authors.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Mühlbacher, W. *et al.* Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun.* **5**:4310 doi: 10.1038/ncomms5310 (2014).