Title
Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH

Journal
Molecular Cell, 59(5)

ISSN
1097-2765

Authors
Luo, Jie
Cimermancic, Peter
Viswanath, Shruthi
et al.

Publication Date
2015-09-01

DOI
10.1016/j.molcel.2015.07.016

Peer reviewed

HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2016 September 03.

Published in final edited form as:

Mol Cell. 2015 September 3; 59(5): 794–806. doi:10.1016/j.molcel.2015.07.016.

Architecture of the human and yeast general transcription and DNA repair factor TFIIH

Jie Luo[1,#], **Peter Cimermancic**[2,#], **Shruthi Viswanath**[2], **Christopher C. Ebmeier**[3], **Bong Kim**[1], **Marine Dehecq**[4,5], **Vishnu Raman**[3], **Charles H. Greenberg**[2], **Riccardo Pellarin**[2], **Andrej Sali**[2], **Dylan J. Taatjes**[3], **Steven Hahn**[4], and **Jeff Ranish**[1,*]

[1]Institute for Systems Biology, 401 Terry Ave N, Seattle, WA 98109, USA

[2]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Sciences, University of California, San Francisco, San Francisco, CA, 94158, USA

[3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80303, USA

[4]Fred Hutchinson Cancer Research Center, Division of Basic Sciences, 1100 Fairview Ave N, PO Box 19024, Mailstop A1-162, Seattle, WA 98109, USA

[5]Génétique des Interactions Macromoléculaires, Institut Pasteur, CNRS UMR3525, 25-28 rue du docteur Roux, 75015 Paris, France

Summary

TFIIH is essential for both RNA polymerase II transcription and DNA repair, and mutations in TFIIH can result in human disease. Here, we determine the molecular architecture of human and yeast TFIIH by an integrative approach using chemical crosslinking/mass spectrometry (CXMS) data, biochemical analyses, and previously published electron microscopy maps. We identified four new conserved “topological regions” that function as hubs for TFIIH assembly and more than 35 conserved topological features within TFIIH, illuminating a network of interactions involved in TFIIH assembly and regulation of its activities. We show that one of these conserved regions, the p62/Tfb1 Anchor region, directly interacts with the DNA helicase subunit XPD/Rad3 in native TFIIH and is required for the integrity and function of TFIIH. We also reveal the structural basis for defects in patients with Xeroderma pigmentosum and Trichothiodystrophy, with mutations found at the interface between the p62 Anchor region and the XPD subunit.

* Corresponding author: J.R. (p) 206-732-1357, jranish@systemsbiology.org Running Title: Architecture of human and yeast TFIIH.

# equal contribution

Author Contributions

J.R. conceived the project. J.R., J.L., A.S., S.H., and D.T. designed the experiments. B.K. and C. E. purified TFIIH, J.L. performed CXMS analyses and data interpretation, P.C., S.V., R.P., and C.G. performed integrative modeling, M.D. performed Tfb1 deletion studies, V.R. performed p62 deletion studies, all authors contributed to writing the manuscript.

Introduction

The conserved, 10-subunit TFIIH complex plays central roles in both RNA polymerase II (Pol II) transcription (Grunberg and Hahn, 2013; Thomas and Chiang, 2006) and DNA repair (Compe and Egly, 2012). TFIIH, the only general transcription factor with enzymatic function, is a component of the Pol II preinitiation complex (PIC). After PIC assembly, the DNA strands around the transcription start site are separated, allowing Pol II access to the template strand followed by transcription initiation, promoter clearance, and elongation. TFIIH plays key roles throughout this process. The TFIIH subunit XPB/Ssl2 (human/yeast subunits) is an ATP-dependent translocase that promotes DNA strand separation and promoter escape (Grunberg et al., 2012; Moreland et al., 1999; Tirode et al., 1999). During the transition from initiation to promoter clearance, the kinase subunit CDK7/Kin28 phosphorylates Ser5 and Ser7 within the Pol II carboxyl terminal domain (CTD). This phosphorylation initiates a cascade of phosphorylation/dephosphorylation events on the CTD that correlates with dissociation of Pol II from the initiation machinery and its association with elongation and mRNA processing factors (Larochelle et al., 2012).

During nucleotide excision repair (NER), TFIIH is required for opening DNA around lesions that disrupt base pairing to permit excision of the damaged DNA and its replacement by a new DNA fragment (Compe and Egly, 2012). The XPB ATPase is required for the stable association of TFIIH with damaged DNA (Oksenych et al., 2009). Opening of the DNA around the lesion is driven by both XPB ATPase function as well as the helicase activity of the TFIIH XPD subunit, resulting in an asymmetric bubble surrounding the DNA lesion (Coin et al., 2007).

TFIIH can be resolved biochemically into a seven subunit core complex containing XPD/Rad3, XPB/Ssl2, p62/Tfb1, p52/Tfb2, p44/Ssl1, p34/Tfb4, and p8/Tfb5 and a three subunit Cdk Activating Kinase (CAK) complex containing CDK7/Kin28, cyclin H/Ccl1, and MAT1/Tfb3. Many studies have demonstrated protein-protein interactions between TFIIH subunits and found that the activities of XPB/Ssl2, XPD/Rad3, and CDK7/Kin28 can be regulated by some of these interactions. For example, p52 interacts directly with XPB and this interaction appears important for stimulation of XPB ATPase activity (Fregoso et al., 2007; Jawhari et al., 2002). XPD interacts directly with p44, which stimulates XPD helicase activity (Coin et al., 1998; Dubaele et al., 2003). XPD/Rad3 also interacts directly with the CAK *via* its MAT1/Tfb3 subunit inhibiting the helicase activity of XPD (Busso et al., 2000; Sandrock and Egly, 2001). In addition, CDK7 activity is regulated by its association with CAK subunits cyclin H and MAT1 (Helenius et al., 2011) as well as by its association with core TFIIH (Rossignol et al., 1997).

Importantly, mutations in TFIIH subunits are associated with many forms of cancer (Manuguerra et al., 2006; Wang et al., 2008), and autosomal recessive disorders, such as Xeroderma Pigmentosum (XP), Tricothiodystrophy (TTD), and the combined symptoms of XP and Cockayne syndrome (XP/CS) (Compe and Egly, 2012). Disease-associated mutations have been found only in XPB, XPD, and p8. Interestingly, most of the XPB and XPD mutations found in patients do not directly affect enzymatic activity, but rather the interactions of these subunits with their regulatory partners (Compe and Egly, 2012).

Structural information on TFIIH has been provided by EM studies of yeast and human TFIIH (Chang and Kornberg, 2000; Gibbons et al., 2012; He et al., 2013; Schultz et al., 2000), X-ray crystallographic studies of human CDK7 and cyclin H and the archaeal homologs of XPB/Ssl2 and XPD/Rad3, as well as X-ray and NMR structures of small domains of subunits p62, Tfb1, p44, and a Tfb2/Tfb5 complex (summarized in (Gibbons et al., 2012)). Recently, a model of a seven subunit core TFIIH was presented based on a combined EM and chemical crosslinking/mass spectrometry (CXMS) study of a yeast PIC in which each TFIIH subunit was represented as 1 or 2 spheres (Murakami et al., 2013), providing the highest resolution structural model of core TFIIH to date. However, the relatively low resolution of this model is insufficient for mapping TFIIH domain interactions, and for a mechanistic understanding of TFIIH function.

To understand how the diverse functions of TFIIH in transcription and DNA repair are regulated, we determined the molecular architecture of both human and yeast TFIIH, by using an integrative approach based on data from our CXMS, genetic and biochemical analyses, as well as data from other sources. Investigation of TFIIH from evolutionarily distant species was critical because it yielded complementary information that allowed identification of conserved topological features, and increased the number of informative crosslinks. We identified four new conserved topological regions and more than 35 conserved topological features within TFIIH. One of these conserved regions, the Anchor region in p62/Tfb1, directly interacts with XPD/Rad3 and is required for the integrity and function of TFIIH. Furthermore, XPD mutations found in patients with XP and TTD localize to the XPD-p62 Anchor region interface, suggesting that an altered XPD-p62 interaction contributes to the pathogenesis of these diseases. Our results provide a structural framework for understanding how TFIIH functions during transcription and DNA repair, as well as insights into the molecular basis for several disease-causing mutations in TFIIH.

Results

Chemical crosslinking and mass spectrometry analysis of human and yeast TFIIH

Human and yeast TFIIH were purified as described in Experimental Procedures (Figs. 1 and S1). We next used a CXMS approach to identify pairs of lysine and/or N-terminal residues that are in spatial proximity to each other in both complexes. CXMS analysis of both human and yeast TFIIH yields important complementary information. Although the orthologs share 21–51% sequence identity, most of the lysine residues are not conserved between the human and yeast subunits; moreover, the orthologs produce predominately unrelated peptides upon trypsin digestion. Human and yeast TFIIH were each crosslinked using the homo-bifunctional, amine-reactive crosslinking reagent bis(sulfosuccinimidyl)suberate (BS3). The samples were then analyzed by MS, and database searching was used to identify crosslinked peptides. Confidently identified crosslinks were used to assemble site-specific linkage maps of all the crosslinked residues between and within the TFIIH subunits (Table S1). Altogether, we identified 89 unique intra-protein crosslinks (a.k.a. intralinks) and 67 unique inter-protein crosslinks (a.k.a. interlinks) from human TFIIH in one experiment (Fig. 1B), and 193 intralinks and 101 interlinks from yeast TFIIH in two independent experiments (Fig. 1D).

To map the positions of these crosslinks, we built comparative models for parts of all subunits, using Modeller 9.9 (Sali and Blundell, 1993), based on template structures identified by HHpred (Soding et al., 2005) and RaptorX (Kallberg et al., 2012). Combined with atomic models from Xray crystallography and NMR spectroscopy, the structural models cover ~47% and ~52% of yeast and human TFIIH sequences, respectively. To evaluate the crosslinks, we mapped them onto the structural models and measured their Cα-Cα distances (Fig. S2A). The BS3 crosslinker has a linker arm of 11.4 Å when fully extended and can crosslink two lysine residues whose Cα atoms are up to 30 Å apart (Merkley et al., 2014). To account for protein dynamics and potential inaccuracies in the comparative models, we used a Cα-Cα distance of 34 Å as the theoretical crosslinking limit. For yeast TFIIH intralinks, 115 out of 193 crosslinks can be mapped onto the structural models (the remainder include at least one residue not present in the models). 108 (~94%) of these crosslinks have Cα-Cα distances less than 34 Å. Seven crosslinks (~6%) had Cα-Cα distances between 36–45 Å, all of which map to putative flexible regions. For human TFIIH intralinks, 71 out of 89 crosslinks can be mapped onto models, and all crosslinks (100%) have Cα-Cα distances less than 34 Å.

To evaluate the interlinks, we mapped crosslinks onto the X-ray structure of Tfb5-Tfb2c (Kainov et al., 2010a), and a model of CDK7/Kin28-cyclin H/Ccl1 based on the CDK12-Cyclin K structure (Bosken et al., 2014). All seven interlinks that could be mapped onto these structures have Cα-Cα distances between 8.5–17.7 Å (Fig. S2B).

Together, these results indicate that our crosslinking data provide useful and reliable distance restraints for modeling the subunit architecture of human and yeast TFIIH. Further support for the quality of our CXMS data comes from the fact that many previously described protein-protein interactions (PPIs) between TFIIH subunits are reflected in our data (Table S2).

Molecular architecture of TFIIH revealed by integrative modeling

We applied an integrative modeling approach (Experimental Procedures and Fig. S3) (Russel et al., 2012) to compute the architecture of the human and yeast TFIIH complexes (Fig. 2). As integrative models are computed from all available data simultaneously, they are often more accurate, precise and complete than those produced by traditional methods. The input data included 421 unique BS3 crosslinks (Table S1), electron microscopy (EM) density maps (Gibbons et al., 2012; He et al., 2013), including the locations of the CAK subcomplex, XPD /Rad3, and XPB/Ssl2 within these EM maps (Gibbons et al., 2012; Murakami et al., 2013), as well as crystallographic structures and comparative models covering 47% and 52% of the yeast and human sequences, respectively (Fig. S4A; Table S5). Atomic level structural information for two interfaces, namely Tfb2 (C-term)-Tfb5 and Ccl1-Kin28, were included in the representation of these subunits.

The molecular architectures of the yeast and human TFIIH complexes were each computed from an ensemble of solutions that satisfied the input data as well as possible (http://salilab.org/tfiih). An ensemble is shown as a localization density map that represents the probability of any volume element being occupied by a given protein (Figs. 2A, C). The sampling procedure was thorough, as indicated by the similarity of solutions obtained from

two independent sets of runs (Figs. S4B and S4C). We validated the two ensembles against the information used to compute them. First, the ensemble of yeast and human solutions satisfied 87% (365 of 421 used for the modeling) and 90% (322 of 356) of the crosslinks, respectively (Fig. S5A). Second, 83% and 100% of the 207 and 106 yeast and human solutions, respectively, satisfied the excluded volume and sequence connectivity restraints and all yeast and human solutions satisfied the EM restraints.

Next, we quantified the precision of the ensemble of solutions. The clustering analysis identified 3 dominant clusters in both ensembles of solutions (Figs. S4B and S4C). The precision of the ensemble is sufficient to pinpoint the locations and orientations of the constituent proteins and even domains (Table S6). XPD and Rad3 were the subunits determined at the highest precision of 9.2 Å and 17.7 Å, respectively (Fig. S5B).

While the overall subunit organizations of the human and yeast cores are similar, their shapes are different. This variation arises from the difference between the EM density maps used for modeling (Fig. 2). The yeast core has a globular shape, while the human core has a rod-like shape, mainly because XPB localizes more centrally in the human than in the yeast model. Due to the different shapes of the two complexes, several differences in the contact maps for human and yeast TFIIH were also observed. First, due to the elongated rod-like structure of human TFIIH, no contacts between XPD and p34, p52, or p8 were observed in the human TFIIH model. In contrast, in the yeast model, Rad3 contacts nearly all of the core subunits (Fig. S5C, blue rectangle). Second, XPB makes extensive contacts with p62 in the human model, whereas only few contacts were observed between Ssl2 and Tfb1 in the yeast model (Fig. S5C, red circle). Interestingly, the human rod-like core structure better fits the data as indicated by (i) significantly less variation among the ensemble of solutions for all the domains of the core subunits of human TFIIH compared to their yeast counterparts (Fig. S5B), (ii) fewer inconsistent crosslinks for the human core TFIIH model than for the yeast core model (Fig. S5A), and (iii) the higher population of the dominant cluster indicating a unique structural solution (Figs. S4B and S4C).

Conserved topological regions in TFIIH

To consolidate the human and yeast CXMS data and better interpret the structural models, we aligned the human and yeast TFIIH subunits based on phylogenetic sequence alignments of the core subunits (Bedez et al., 2013) and direct sequence alignments of the CAK subunits (Table S3). The crosslinks were mapped onto the alignments (Fig. 3A), allowing us to directly compare the crosslinks from human and yeast TFIIH complexes. Identification of crosslinks between positions close in the alignment of the human and yeast TFIIH sequences not only provides additional confidence in the data, but also suggests evolutionarily conserved spatial organization of the crosslinked residues within TFIIH.

Because the spatial distributions of the lysine residues in the human and yeast orthologs are different and BS3 can crosslink lysine residues whose Cα atoms are 5–30 Å apart, we reasoned that for the purpose of identifying conserved structural features, the exact positions of the crosslinked lysine residues would be less important than the domains in which the crosslinked residues reside. That is, the conserved crosslinks reflect the conserved proximity of the associated domains. We mapped the positions of crosslinked residues onto the aligned

sequences of human and yeast TFIIH subunits (Figs. 3A and S6A) and then identified the conserved crosslinks between domains (Fig. 3B and Table S2). Not surprisingly, many conserved crosslinks map to structured domains suggesting direct domain-domain interactions; previous PPI studies among TFIIH subunits provide support for many of these interactions (Table S2 and below).

Approximate sequence positions are conserved for 75 (74%) and 46 (69%) of the 101 yeast and 67 human interlinks, respectively. Thus, these interlinks reveal evolutionarily conserved sequence segments (i.e., “topological regions”) and proximities between them. Interestingly, some of the conserved crosslinks map to sequence segments that lack Pfam domain annotation, suggesting such segments could have important roles in the organization of TFIIH subunits even though they are not conserved in sequence. In addition, we have also identified segments that contain multiple Pfam domains but with conserved crosslink patterns, suggesting similar roles for some of the known domains in the spatial organization of the TFIIH subunits. Thus, to expand our knowledge of structurally and functionally important regions in TFIIH, we introduce four new topological regions, termed the Anchor (p62/Tfb1), the Hub (p52/Tfb2), the Lock (XPB/Ssl2), and the Latch (MAT1/Tfb3) (Fig. 3A, Tables S2 and S4). The interlinks that map to these four topological regions account for 48% of the conserved crosslinks for yeast TFIIH and 70% of the conserved crosslinks for human TFIIH. As expected, these topological regions are also indicated by their relative contact frequencies in our model ensembles, which define how often any pair of residues contact each other (within a 5 Å cutoff) in a cluster of solutions (Alber et al., 2007)(Fig. S5C and below). We next discuss several new structural features of TFIIH revealed by the crosslinking data and the models.

p44/Ssl1-p34/Tfb4 dimer forms the base of the TFIIH core

Both p34/Tfb4 and p44/Ssl1 have an N-terminal VWA domain and a C-terminal Zn RING motif (Kellenberger et al., 2005; Schmitt et al., 2014; Whittaker and Hynes, 2002). The C-terminal RING motif of p44 interacts with the VWA domain of p34 (Fribourg et al., 2001; Kellenberger et al., 2005; Schmitt et al., 2014), and the VWA domain of p44 interacts with p34 (Iyer et al., 1996). In agreement with these studies, the proximity of p44/Ssl1 and p34/Tfb4, as well as their respective VWA domains is predicted by our models (Figs. 2 and S5C). We found an extensive network of crosslinks between these two subunits including three crosslinks from yeast TFIIH and five crosslinks from human TFIIH between the two VWA domains of these subunits (Figs. 3 and S6B, Table S2). VWA domains are known to be involved in protein-protein interactions and dimerization (Becker et al., 2014; Springer, 2006). Conserved crosslinks between the VWA domains of p34/Tfb4 and p44/Ssl1 strongly suggest that these domains interact directly to form a heterodimer in TFIIH. Further, we identified three crosslinks between the VWA domain of Tfb4 and the Ssl1 RING motif in agreement with previous findings (Fribourg et al., 2001). The structurally homologous p34/Tfb4 and p44/Ssl1 proteins are the only proteins known that contain both a VWA domain and Zn RING motif (Whittaker and Hynes, 2002), and their homologs exist in all analyzed eukaryotes (Bedez et al., 2013). In light of these observations, we hypothesize that p34/Tfb4 and p44/Ssl1 evolved *via* gene duplication soon after the emergence of eukaryotes and the ancestral homologs homo-dimerized through VWA-VWA interactions.

Our models also predict that p44/Ssl1 and p34/Tfb4 are frequently in contact with the other core subunits (Fig. S5C, green squares). We identified many conserved crosslinks between p44/Ssl1 and p34/Tfb4 and the other core TFIIH subunits (Fig. 3 and below). Thus, through VWA domain dimerization, we propose that a p34/Tfb4-p44/Ssl1 dimer forms the base of core TFIIH that interacts with all core subunits.

The p62/Tfb1 Anchor region

p62/Tfb1 has an N-terminal pleckstrin homology (PH) domain, followed by two BSD domains (BSD1 and BSD2) and a putative 3-helix bundle motif at its C-terminus (Fig. 3). The PH domain interacts with the XPG endonuclease, the C-terminus of the large subunit of TFIIE, and the second transactivation domain of the p53 tumor suppressor protein (Di Lello et al., 2008; Gervais et al., 2004; Okuda et al., 2008). The PH domain also interacts with XPD and XPB (Iyer et al., 1996). The BSD domains are predicted to form three-helical bundles (Doerks et al., 2002). We identified crosslinks between the PH domain and (i) the VWA domain of p44/Ssl1, (ii) the p34 VWA domain and the Tfb4 RING finger and (iii) the HD1a/FeS region of Rad3 (Table S2, Fig. 4). We also identified conserved intralinks between the PH domain and the BSD domains. The C-terminal helical bundle crosslinks to BSD1 in both human (Fig. S6A) and yeast TFIIH (Murakami et al., 2013), suggesting that BSD1, BSD2 and the C-terminal domain could interact to form a large helical bundle.

A large region between BSD2 and the C-terminal helical bundle (residues 254–457 for p62 and 332–546 for Tfb1) crosslinks extensively to the VWA domains of p34/Tfb4 and p44/Ssl1 and the HD1a-4FeS and HD2 domains of XPD/Rad3 (Table S2, Fig. 4). This region, which we call the Anchor region, possesses no obvious structural motifs and little sequence conservation. Our models also predict that the Anchor region contacts the 4FeS domain and the HD2/C-terminus of XPD/Rad3, as well as the p44/Ssl1 and p34/Tfb4 subunits (Fig. S5C, red squares). There is previous support for Anchor region interaction with XPD and Rad3 (Jawhari et al., 2004; Murakami et al., 2013) as well as Ssl1 (Murakami et al., 2013), but not for Anchor region interaction with p44, p34 or Tfb4. The lack of identified crosslinks between the Anchor region and the other domains within p62/Tfb1 suggest that the Anchor region protrudes from the rest of p62/Tfb1 to allow it to interact with other subunits.

The Hub region of p52/Tfb2 connects p52/Tfb2 and XPB/Ssl2 to the base

The structure of the N-terminus of p52/Tfb2 is predicted to be similar to the HEAT repeats of human transportin 3 (Maertens et al., 2014). The C-terminus of p52/Tfb2 contains a domain that can dimerize with itself or with p8/Tfb5 (Kainov et al., 2008). A region before the dimerization domain, referred to as the hA-D domain (Fig. 3A), is predicted by RaptorX to be an independent domain containing four alpha helices but with no strong similarity to known structures. Since similar linkages were observed for both the C-terminus and the hA-D domains, we propose that these two domains function as one structural unit, which we call the Hub region (residues 301–462 for p52 and 333–513 for Tfb2). The Hub region contains 3 of the 5 highly conserved stretches of amino acids in p52 (Jawhari et al., 2002). The interaction between the p52/Tfb2 dimerization domain and p8/Tfb5 is supported by our models and the identification of conserved crosslinks between these two subunits/domains (Fig. S6D) whose Cα-Cα distances are well within the theoretical distance for BS3 (Fig.

S2B). Besides p8/Tfb5 interactions, many other conserved crosslinks were identified between the Hub region and other TFIIH subunits (Table S1 and S2, Fig. S6B). The Hub region crosslinks extensively to the VWA domains of p44/Ssl1 and p34/Tfb4. These conserved crosslinks indicate that p52/Tfb2 is anchored to core TFIIH via interactions between the Hub region and the VWA domains of p44/Ssl1 and p34/Tfb4 subunits.

Conserved crosslinks were also identified between the p52/Tfb2 Hub region and the C-terminus of XPB/Ssl2 as well as four additional crosslinks to other regions of XPB in human TFIIH (Fig. S6D, Table S2). Amino acids 1–135 and 304–381 of p52 are important for interaction with XPB (Fregoso et al., 2007; Jawhari et al., 2002), and amino acids 44–208 of XPB are important for interaction with p52 (Jawhari et al., 2002). Our observation that the p52 Hub region crosslinks to the N-terminus of XPB is in agreement with these findings. However, no crosslinks were identified between the N-terminal HEAT repeats of p52/Tfb2 and XPB/Ssl2. Furthermore, no interaction between the C-terminus of XPB and p52 has been reported. We also find that p8/Tfb5 crosslinks to both the N-terminus and the C-terminus of XPB/Ssl2 and some of the sites of crosslinking on XPB/Ssl2 overlap with those between XPB/Ssl2 and the Hub region (Fig. S6D), indicating that p8/Tfb5 also directly interacts with XPB/Ssl2. Further support for a p8/Tfb5-XPB/Ssl2 interaction comes from TFIIH co-immunoprecipitation (Co-IP) studies in which Ssl2 levels are preferentially affected in the absence of Tfb5 compared to the other subunits (Fig. S7B). In agreement with previous findings that p52 is needed to anchor XPB to the TFIIH complex (Jawhari et al., 2002), our crosslinking data show that the Hub region of p52/Tfb2, together with its interaction partner p8/Tfb5, interacts with both the N and C-terminal parts of XPB/Ssl2 in TFIIH. The proposal that the Hub subdomains, Hub-A (hA-D) and Hub-B (dimerization), function as a single unit is supported by their similar contact frequency profiles in our models (Fig. S5C, cyan squares). Together, the data indicate that the Hub region, along with p8/TFB5, anchors XPB/Ssl2 to the base of TFIIH through its interaction with the VWA domains of p44/Ssl1 and p34/Tfb4.

The Lock region of XPB/Ssl2

The *Archaeoglobus fulgidus* XPB/Ssl2 homolog (*Af*XPB) contains DRD, HD1 and HD2 domains (Fan et al., 2006), while the eukaryotic proteins also contain N- and C-terminal extensions. The N-terminal extension contains a conserved region identified as a putative DNA binding domain (*DBD*) (Jawhari et al., 2002), which we predict has a fold that is similar to the fold found in the p52/Tfb2 dimerization domain. The C-terminal extension is rich in acidic residues (DE-rich region). Several lines of evidence suggest that the N- and C-terminal extensions form a structural unit that regulates the function(s) of XPB/Ssl2, which we called the Lock region. Lock-N (residues 31–226 for XPB and 88–285 for Ssl2) and Lock-C (residues 669–782 for XPB and 715–843 for Ssl2) refer to the N- and C-terminal extensions respectively. First, we identified conserved intralinks between the Lock-N and Lock-C segments. Second, the C-terminus of p8 crosslinks to both Lock-N and Lock-C (Figs. S6B and S6D). Third, the p52 Hub region crosslinks to Lock-C and to the DRD domain which in turn crosslinks to Lock-N. These data indicate that the Lock-C segment is in close proximity to the Lock-N segment.

Besides the interactions with p52/Tfb2 and p8/Tfb5 discussed above, the Lock-N segment crosslinks to the HD2 domain of XPD, and to the N-terminus of p44/Ssl1 and the MAT1/Tfb3 Latch region (discussed below; Table S2, Fig. S6C). These results are supported by previous studies showing that the N-terminal region of XPB interacts with XPD and MAT1 (Busso et al., 2000; Iyer et al., 1996), and the N-terminal region of Ssl2 interacts with the N-terminus of Ssl1 and the Tfb3 Latch region (Murakami et al., 2013). The Lock-N segment also interacts with the XPG endonuclease (Iyer et al., 1996). Furthermore, we found that Tfb6, a yeast TFIIH-specific subunit, which facilities dissociation of Ssl2 after transcription initiation (Murakami et al., 2012), crosslinks to the Lock-N segment (data not shown). Together, the data indicate that the N- and C-terminal regions of XPB are in close proximity in TFIIH and form a region that interacts with a number of TFIIH subunits including p8/Tfb5, p52/Tfb2, p44/Ssl1 and MAT1/Tfb3, as well as NER factor XPG. Our models also predict that the Lock-N and Lock-C of XPB/Ssl2 are frequently in contact with each other and they have similar interaction profiles (Fig. S5C, black squares). These interactions with the Lock N-C region may anchor XPB to TFIIH and/or regulate its enzymatic activities.

The original proposal for open and closed conformational states of XPB/Ssl2 is based on structural studies of *Af*XPB where an open conformation was observed (Fan et al., 2006). The authors proposed that upon DNA binding HD2 rotates at the glycine-containing hinge by ~170°, bringing the seven helicase motifs together at the cleft formed between HD1 and HD2 to form the closed conformation, as seen in some DNA bound helicases. The open-to-closed conformational change was proposed as the mechanism for activation of XPB DNA strand separation activity. To determine if the CXMS data could distinguish between the two conformations, we mapped the crosslinks onto models of the open and closed conformations of human XPB (Fig. S6E). All three interlinks (between the C-terminus of p8/Tfb5 and XPB/Ssl2: p8:K71-XPB:K476, p8:K71-XPB:K526 and Tfb5:K60-Ssl2:K721) and both intralinks (between Ssl2:K523-K791 and Ssl2:K523-K794) are in close proximity only in the closed form (Fig. S6E). This observation indicates that at least some proportion of XPB/Ssl2 ATPase domains assume the closed orientation in free TFIIH, although a possibility of a mixture of the two states is not excluded.

The MAT1/Tfb3 Latch region links the CDK module to the core

MAT1/Tfb3 has an N-terminal C3HC4 RING finger domain, a central coiled-coil domain with a long helical fibrinogen-like structure, and a hydrophobic region at the C-terminus (Inamoto et al., 1997; Yee et al., 1995). The coiled-coil domain interacts with XPD and XPB and the hydrophobic domain interacts with CDK7/Cyclin H and stimulates CDK7 kinase activity (Busso et al., 2000). We identified many interlinks involving the central coiled-coil domain and the Cterminal hydrophobic domain of MAT1/Tfb3 (Fig. S6C, Table S2) which leads us to propose a new structural motif, termed the Latch region, containing these two domains (residues 109–305 for MAT and 120–314 for Tfb3). Conserved crosslinks were identified between the Latch region and the Lock-N region of XPB/Ssl2, as well as between the Latch region and the HD2 domain of XPD/Rad3. These results are supported by previous results showing that the Latch region interacts with both XPD/Rad3 and XPB/Ssl2 (Busso et al., 2000). Abdulrahman *et al*. reported that the Arch domain of XPD is important for recruitment of the CAK complex because mutations in the Arch reduce the ability of

MAT1 and the CAK to associate with XPD (Abdulrahman et al., 2013). In agreement with this interaction, we identified crosslinks between the Latch and the Arch domain in Rad3 (Fig. S6C). The Latch region also crosslinks to the N-terminus of Ssl1, as well as to Kin28 and Ccl1. In our modeled TFIIH structure, The Latch region of MAT1/Tfb3 contacts multiple domains in XPB/Ssl2 and XPD/Rad3 (Fig. S5C, yellow squares). Taken together, the data indicate that the MAT1/Tfb3 Latch region anchors the CAK module to TFIIH via interactions with XPD/Rad3, XPB/Ssl2 and p44/Ssl1.

The Anchor region of p62/Tfb1 is essential for TFIIH structural integrity

To further understand the function of the domains in Tfb1, we tested growth phenotypes and UV sensitivity of cells containing a series of *TFB1* derivatives as the sole source of Tfb1 (Fig. 5A). Although the PH domain interacts with TFIIE, XPG and other TFIIH subunits, deletion of this domain had no effect on growth or UV sensitivity. Deletion of the BSD1 domain caused mild temperature and UV sensitivity, and deletion of the BSD2 domain had no phenotype. Deletion of the C-terminal helical bundle caused pronounced temperature and UV sensitivity consistent with results of a previous study (Matsui et al., 1995). Most of the conserved crosslinks in p62/Tfb1 were observed in the Anchor region. Strikingly, deletion of the unstructured region in the middle of the Anchor region (Δ401–491) resulted in a lethal phenotype, showing that this region is essential for the function of TFIIH.

We next immunoprecipitated Tfb1 derivatives to determine which deletions affect the integrity of TFIIH (Figs. 5B and S7A). Although deletions that remove the BSD1 or BSD2 domains (Δ5–7) did not affect yeast growth rate, they all showed significant reductions in association with Rad3, Tfb3, Ssl2 and Kin28. Deletions removing portions of the Anchor region (Δ9–11) showed progressively more severe defects in subunit interactions as deletions progressed from the N to C-terminus. Δ9–10 show similar defects as Δ5–7 with an additional defect in association with Tfb2. Deletion Δ11, removing the C-terminus of the Anchor region, is defective in all subunit interactions tested. Finally, deletion of the Tfb1 3-helix bundle C-terminus is defective in interactions with Tfb2, Ssl1, and Tfb4. In addition to causing reduced association with the above TFIIH subunits, several of the deletions resulted in increased association with other TFIIH subunits (Fig. 5B). Although the magnitude of this increased association varied in independent experiments, the association of these subunits was consistently higher compared to cells containing wild type Tfb1.

Based on our crosslinking, Tfb1 IP results, and modeling of TFIIH subunit association, there are two groups of Tfb1 deletions: (i) deletions that directly eliminate protein-protein contacts and (ii) deletions that indirectly affect TFIIH subunit association. Mutations within the Anchor likely directly eliminate interactions with Rad3 and/or Ssl1/Tfb4. Mutations that weaken Rad3 interaction are expected to also weaken co-IP of Tfb1 with the CAK subunits and with the interacting subunit Ssl2 as all these subunits bind TFIIH in part via Rad3. Although the Anchor is not predicted to directly contact Tfb2, deletions Δ9–10 likely alter interactions with the Tfb4-Ssl1 heterodimer leading to instability of its associated subunit Tfb2. Removal of the Anchor region C-terminus (Δ11) disrupts interactions of Tfb1 with Tfb4-Ssl1 and Rad3 leading to the failure of Tfb1 to associate with all tested TFIIH subunits. The essential role of the Anchor region in maintaining structural integrity of TFIIH

was confirmed in human cells (Fig. 5C). Deletion of the Anchor region in p62 (residues 328 – 432) prevented co-purification of XPD and XPB, in agreement with findings in the yeast system. MS analysis of FLAG-purified p62 Anchor domain deletion mutant confirmed these results (Table S7).

Deletions Δ5–7 within the BSD1/2 domains are selectively defective in interactions with Rad3 and associated subunits. We speculate that this is due to an indirect effect on the Tfb1 Anchor region since no crosslinks were identified between the BSD domains and Rad3 or its associated subunits. Similarly, a deletion in the C-terminus of the Tfb1 3-helix bundle (Δ13) is defective for association with Ssl1/Tfb4 and its associated subunit Tfb2, yet no crosslinks were identified between the 3-helix bundle and these subunits. We speculate that this deletion alters the conformation of the Anchor region leading to defective association with these three TFIIH subunits. Indirect effects are also likely responsible for the increased association of some subunits with Tfb1 derivatives during IP. For example, deletions within BSD1/2 weaken interaction of Tfb1 with Rad3 while simultaneously strengthening the interactions with Ssl1/Tfb4. Likewise, deletion at the N-terminus of the 3-helix bundle (Δ12) increases binding to Rad3 and associated subunits, likely via an indirect effect mediated though the Anchor region.

Discussion

Architecture of TFIIH

We have summarized the domain-domain proximities in a schematic model of TFIIH, highlighting the identified conserved topological features (Fig. 6). Our model positions p34/Tfb4 and p44/Ssl1 at the base of TFIIH with their VWA domains forming a dimerization interface. The essential Anchor region of p62/Tfb1 interacts with both the VWA domains of p34/Tfb4 and p44/Ssl1, and with XPD/Rad3 to anchor XPD/Rad3 to the base of TFIIH. The Hub region of p52/Tfb2, together with p8/Tfb5, interacts with the Lock N-C region of XPB/Ssl2, thus tethering XPB/Ssl2 to the core domain. The Lock region is proposed to be the main regulatory region of XPB/Ssl2 activity *via* its interactions with other TFIIH subunits as well as other factors. The Latch region of MAT1/Tfb3 anchors the CAK module to the TFIIH core through an extensive network of interactions involving the Arch and HD2/CTE domains of XPD/Rad3, the Lock-N region of XPB/Ssl2, and the N-terminal extension of p44/Ssl1. Our model of TFIIH architecture not only unifies many previous findings, but also reveals new insights into the structure and function of the 10-subunit TFIIH complex. For example, the organization of some of the core TFIIH subunits is different from a previously published model of core TFIIH in which p62/Tfb1 is not positioned adjacent to XPD/Rad3, and p44/Ssl1 is not positioned adjacent to p34/Tfb4 (Fuss and Tainer, 2011). Compared to the model presented by Murakami et. al., the resolution of our model is higher but the subunit organization is quite similar, except for the orientation of Ssl2, whose C-terminus is positioned closer to p52/Tfb2-p8/Tfb5 in our model (Fig. S4D).

While the overall subunit organization and domain-domain contacts for the human and yeast cores are similar, their shapes as revealed by the respective electron microscopy studies are quite different. The human core TFIIH has a more elongated shape compared to the yeast core TFIIH, mainly due to more central localization of XPB. Both EM maps were obtained

from a preparation of holo-TFIIH (Gibbons et al., 2012; He et al., 2013) however, the differences could arise as a result of different isolation and purification protocols, from collection and processing of micrographs, and single-particle reconstructions. Alternatively, the different shapes might reflect species-specific differences in the architecture of the complexes.

An essential p62/Tfb1 Anchor region links XPD/Rad3 to TFIIH

The identification of the essential and conserved Anchor region in p62/Tfb1 provides new insight into how XPD/Rad3 is linked to the core at relatively high resolution (precision of 10.1 Å RMSD)(Figs. 7 and S5B). Previous studies implicated p44/Ssl1 as a subunit that links XPD to the core (Abdulrahman et al., 2013). Our results show that the Anchor region interacts with XPD/Rad3 and is required for association of XPD/Rad3 with p62/Tfb1. Our data also suggest that p62/Tfb1 regulates the enzymatic activities of XPD/Rad3 (below). The Anchor region is poorly conserved at the sequence level, yet it is a conserved topological feature in TFIIH. Furthermore, the Anchor region is predicted to be intrinsically disordered. Intrinsically disordered regions (IDRs) are often found in large protein complexes where they can function as versatile protein-protein interaction surfaces, in part because by being disordered, they can sample larger surface areas (Tompa, 2012).

Insights into the molecular basis for TFIIH-related diseases

The high precision of the TFIIH model allowed us to place the atomic model of XPD into the density map and to characterize the interface between XPD, p62, and p44 in more detail (Fig. 7). Specifically, we mapped the positions of XPD mutations found in patients with XP/TTD or XP/CS. Three mutations, R616P, D673G, and G675R, found in patients with XP/TTD, TTD, and XP/CS respectively, are located at the interface between XPD and the p62 Anchor region. The R616P and G675R mutations prevent XPD from interacting with p44 in immunoprecipitation assays (Dubaele et al, 2003), and the D673G mutation is also thought to be associated with TFIIH destabilization (Fan et al., 2008). Interestingly, our model suggests that p44 is in close contact with the C-terminus of XPD, while these three mutations are at the interface between XPD and the Anchor region of p62, consistent with the importance of the Anchor region in securing XPD to the core. Our results suggest that a compromised XPD-p62 Anchor region interaction contributes to impaired TFIIH function in patients carrying these mutations. In support of this conclusion, XPD variants containing the R616P and G675R mutations failed to co-purify with TFIIH under stringent purification conditions (Dubaele et al., 2003). The defective interaction was previously attributed to an impaired p44-XPD interaction; however, our results suggest that this defect is an indirect effect and that an impaired p62 Anchor region-XPD interaction contributes to the disease phenotypes in these patients.

Our findings also provide insight into the molecular basis for TTD in patients carrying mutations in p8. The p8 subunit has been proposed to stabilize TFIIH, as reduced levels of XPB are observed in cells upon knockdown (Coin et al., 2006) or mutation of p8 (Giglia-Mari et al., 2004), whereas overexpression of p8 restores normal levels of XPB (Aguilar-Fuentes et al., 2008; Coin et al., 2006). The p8 subunit also stimulates the XPB ATPase, yet no direct interaction between these proteins was known (Coin et al., 2006). In addition, Tfb5

(yeast p8 ortholog) variants with TTD-associated mutations have reduced affinity for Tfb2 (yeast p52 ortholog)(Kainov et al., 2010b). Our models, based on extensive crosslinks between p8/Tfb5, p52/Tfb2, and XPB/Ssl2, indicate that these proteins form an intricate interaction network.

Previous findings showed that p52 is needed to anchor XPB to the TFIIH complex (Jawhari et al, 2002). Our model indicates that the Hub region of p52/Tfb2, together with its interaction partner p8/Tfb5, interacts with multiple domains within XPB/Ssl2, including its N- and C-termini. Thus, it is likely that both p8 and p52 stabilize XPB and regulate its ATPase activity *via* direct interactions within TFIIH. In TTD-A, an impaired p8-p52 interaction is predicted to disrupt the p8-p52-XPB interaction network. This disruption impacts TFIIH function by destabilizing XPB and/or by reducing its ATPase activity. In either scenario, disease-causing mutations in p8 would be expected to result in defects in transcription and DNA repair (Aguilar-Fuentes et al., 2008; Coin et al., 2006; Giglia-Mari et al., 2004; Ranish et al., 2004).

Experimental Procedures

See Supplementary Experimental Procedures for details on strains and plasmids, methods of protein purification, IP and yeast growth assays.

Determination of TFIIH architecture by integrative modeling

Our integrative approach to determining the architecture of TFIIH proceeds through four stages (Alber et al., 2007; Russel et al., 2012) (Fig. S3): (1) gathering of data, which includes crosslinking data and electron microscopy maps, as well as previously published data from the literature and public databases, (2) representation of subunits, defined based on input information (eg, atomic and coarse-grain sphere representation for parts with known and unknown structures, respectively), and simultaneous translation of all of the data into spatial restraints and a single scoring function that ranks alternative models, (3) configurational sampling to produce an ensemble of models that satisfies the restraints, and (4) analysis of the ensemble. The modeling protocol (*ie*, stages 2, 3, and 4) was scripted using the IMP *Python Modeling Interface* (IMP-PMI; https://github.com/salilab/pmi), version c522e0abc, a library to model macromolecular complexes based on our open source *Integrative Modeling Platform* package (http://salilab.org/imp/), version c93138f (Russel et al., 2012). All the input, output, and protocol files are available at http://salilab.org/tfiih. See Supplemental Experimental Procedures for details.

Crosslinking and mass spectrometry analysis

~100 pmoles of purified yeast TFIIH or human TFIIH were crosslinked by addition of BS3 (Thermo Scientific; 100 mM in 20 mM triethanolamine, pH 8) to 5 mM for 30 minutes at 25º C. Samples were digested with trypsin, and the resulting peptides were fractionated by SCX chromatography and analyzed by MS. The crosslinked peptides were identified by searching the MS data against appropriate databases using two different algorithms: pLink and in-house designed Nexus (in preparation; J.L. and J.R.; short description in Knutson et al., 2014). A 5% of FDR was used for both pLink and Nexus searches. Details are provided in Supplemental Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We wish to thank James Fishburn for assay of yeast TFIIH transcription activity, and L. Warfield for help with *in vivo* TFIIH analysis. We thank JM Egly for providing the human XPD antibody. This work was supported by grants from the US National Institutes of Health (NIH)(P50 GM076547 to J.R., R21CA175849 to J.R. and D.J.T, RO1 GM053451 to S.H., as well as P41 GM109824 and R01 GM083960 to A.S.), and the National Science Foundation (NSF) (MCB 1244175 to D.J.T.), as well as by a Howard Hughes Medical Institute Predoctoral Fellowship to P.C. M.D. was a student in the Magistère de Génétique Graduate Program at Université Paris Diderot, Sorbonne Paris Cité.

References

Abdulrahman W, Iltis I, Radu L, Braun C, Maglott-Roth A, Giraudon C, Egly JM, Poterszman A. ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. Proc Natl Acad Sci U S A. 2013; 110:E633–E642. [PubMed: 23382212]

Aguilar-Fuentes J, Fregoso M, Herrera M, Reynaud E, Braun C, Egly JM, Zurita M. p8/TTDA overexpression enhances UV-irradiation resistance and suppresses TFIIH mutations in a Drosophila trichothiodystrophy model. PLoS Genet. 2008; 4:e1000253. [PubMed: 19008953]

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al. Determining the architectures of macromolecular assemblies. Nature. 2007; 450:683–694. [PubMed: 18046405]

Becker AK, Mikolajek H, Paulsson M, Wagener R, Werner JM. A structure of a collagen VI VWA domain displays N and C termini at opposite sides of the protein. Structure. 2014; 22:199–208. [PubMed: 24332716]

Bedez F, Linard B, Brochet X, Ripp R, Thompson JD, Moras D, Lecompte O, Poch O. Functional insights into the core-TFIIH from a comparative survey. Genomics. 2013; 101:178–186. [PubMed: 23147676]

Bosken CA, Farnung L, Hintermair C, Merzel Schachter M, Vogel-Bachmayr K, Blazek D, Anand K, Fisher RP, Eick D, Geyer M. The structure and substrate specificity of human Cdk12/Cyclin K. Nature communications. 2014; 5:3505.

Busso D, Keriel A, Sandrock B, Poterszman A, Gileadi O, Egly JM. Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. J Biol Chem. 2000; 275:22815–22823. [PubMed: 10801852]

Chang WH, Kornberg RD. Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. Cell. 2000; 102:609–613. [PubMed: 11007479]

Coin F, Marinoni JC, Rodolfo C, Fribourg S, Pedrini AM, Egly JM. Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. Nature genetics. 1998; 20:184–188. [PubMed: 9771713]

Coin F, Oksenych V, Egly JM. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell. 2007; 26:245–256. [PubMed: 17466626]

Coin F, Proietti De Santis L, Nardo T, Zlobinskaya O, Stefanini M, Egly JM. p8/TTD-A as a repair-specific TFIIH subunit. Mol Cell. 2006; 21:215–226. [PubMed: 16427011]

Compe E, Egly JM. TFIIH: when transcription met DNA repair. Nat Rev Mol Cell Biol. 2012; 13:343–354. [PubMed: 22572993]

Di Lello P, Miller Jenkins LM, Mas C, Langlois C, Malitskaya E, Fradet-Turcotte A, Archambault J, Legault P, Omichinski JG. p53 and TFIIEalpha share a common binding site on the Tfb1/p62 subunit of TFIIH. Proc Natl Acad Sci U S A. 2008; 105:106–111. [PubMed: 18160537]

Doerks T, Huber S, Buchner E, Bork P. BSD: a novel domain in transcription factors and synapse-associated proteins. Trends in biochemical sciences. 2002; 27:168–170. [PubMed: 11943536]

Dubaele S, Proietti De Santis L, Bienstock RJ, Keriel A, Stefanini M, Van Houten B, Egly JM. Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. Mol Cell. 2003; 11:1635–1646. [PubMed: 12820975]

Fan L, Arvai AS, Cooper PK, Iwai S, Hanaoka F, Tainer JA. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Mol Cell. 2006; 22:27–37. [PubMed: 16600867]

Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, Roberts VA, Cooper PK, Tainer JA. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell. 2008; 133:789–800. [PubMed: 18510924]

Fregoso M, Laine JP, Aguilar-Fuentes J, Mocquet V, Reynaud E, Coin F, Egly JM, Zurita M. DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility. Molecular and cellular biology. 2007; 27:3640–3650. [PubMed: 17339330]

Fribourg S, Romier C, Werten S, Gangloff YG, Poterszman A, Moras D. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. Journal of molecular biology. 2001; 306:363–373. [PubMed: 11237605]

Fuss JO, Tainer JA. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. DNA repair. 2011; 10:697–713. [PubMed: 21571596]

Gervais V, Lamour V, Jawhari A, Frindel F, Wasielewski E, Dubaele S, Egly JM, Thierry JC, Kieffer B, Poterszman A. TFIIH contains a PH domain involved in DNA nucleotide excision repair. Nat Struct Mol Biol. 2004; 11:616–622. [PubMed: 15195146]

Gibbons BJ, Brignole EJ, Azubel M, Murakami K, Voss NR, Bushnell DA, Asturias FJ, Kornberg RD. Subunit architecture of general transcription factor TFIIH. Proc Natl Acad Sci U S A. 2012; 109:1949–1954. [PubMed: 22308316]

Giglia-Mari G, Coin F, Ranish JA, Hoogstraten D, Theil A, Wijgers N, Jaspers NG, Raams A, Argentini M, van der Spek PJ, et al. A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat Genet. 2004; 36:714–719. [PubMed: 15220921]

Grunberg S, Hahn S. Structural insights into transcription initiation by RNA polymerase II. Trends in biochemical sciences. 2013; 38:603–611. [PubMed: 24120742]

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. 2012; 19:788–796. [PubMed: 22751016]

He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. Nature. 2013; 495:481–486. [PubMed: 23446344]

Helenius K, Yang Y, Tselykh TV, Pessa HK, Frilander MJ, Makela TP. Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. Nucleic acids research. 2011; 39:5025–5035. [PubMed: 21385826]

Inamoto S, Segil N, Pan ZQ, Kimura M, Roeder RG. The cyclin-dependent kinase-activating kinase (CAK) assembly factor, MAT1, targets and enhances CAK activity on the POU domains of octamer transcription factors. J Biol Chem. 1997; 272:29852–29858. [PubMed: 9368058]

Iyer N, Reagan MS, Wu KJ, Canagarajah B, Friedberg EC. Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein. Biochemistry. 1996; 35:2157–2167. [PubMed: 8652557]

Jawhari A, Boussert S, Lamour V, Atkinson RA, Kieffer B, Poch O, Potier N, van Dorsselaer A, Moras D, Poterszman A. Domain architecture of the p62 subunit from the human transcription/repair factor TFIIH deduced by limited proteolysis and mass spectrometry analysis. Biochemistry. 2004; 43:14420–14430. [PubMed: 15533047]

Jawhari A, Laine JP, Dubaele S, Lamour V, Poterszman A, Coin F, Moras D, Egly JM. p52 Mediates XPB function within the transcription/repair factor TFIIH. J Biol Chem. 2002; 277:31761–31767. [PubMed: 12080057]

Kainov DE, Cura V, Vitorino M, Nierengarten H, Poussin P, Kieffer B, Cavarelli J, Poterszman A. Structure determination of the minimal complex between Tfb5 and Tfb2, two subunits of the yeast transcription/DNA-repair factor TFIIH: a retrospective study. Acta crystallographica. Section D, Biological crystallography. 2010a; 66:745–755.

Kainov DE, Selth LA, Svejstrup JQ, Egly JM, Poterzsman A. Interacting partners of the Tfb2 subunit from yeast TFIIH. DNA repair. 2010b; 9:33–39. [PubMed: 19897425]

Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly JM. Structural basis for group A trichothiodystrophy. Nat Struct Mol Biol. 2008; 15:980–984. [PubMed: 19172752]

Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nature protocols. 2012; 7:1511–1522. [PubMed: 22814390]

Kellenberger E, Dominguez C, Fribourg S, Wasielewski E, Moras D, Poterszman A, Boelens R, Kieffer B. Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. J Biol Chem. 2005; 280:20785–20792. [PubMed: 15790571]

Larochelle S, Amat R, Glover-Cutter K, Sanso M, Zhang C, Allen JJ, Shokat KM, Bentley DL, Fisher RP. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. Nat Struct Mol Biol. 2012; 19:1108–1115. [PubMed: 23064645]

Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Oztop I, Lee K, Pye VE, Cosnefroy O, Snijders AP, et al. Structural basis for nuclear import of splicing factors by human Transportin 3. Proc Natl Acad Sci U S A. 2014; 111:2728–2733. [PubMed: 24449914]

Manuguerra M, Saletta F, Karagas MR, Berwick M, Veglia F, Vineis P, Matullo G. XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. American journal of epidemiology. 2006; 164:297–302. [PubMed: 16707649]

Matsui P, DePaulo J, Buratowski S. An interaction between the Tfb1 and Ssl1 subunits of yeast TFIIH correlates with DNA repair activity. Nucleic acids research. 1995; 23:767–772. [PubMed: 7708491]

Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein science : a publication of the Protein Society. 2014; 23:747–759. [PubMed: 24639379]

Moreland RJ, Tirode F, Yan Q, Conaway JW, Egly JM, Conaway RC. A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. J Biol Chem. 1999; 274:22127–22130. [PubMed: 10428772]

Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. Science. 2013; 342:1238724. [PubMed: 24072820]

Murakami K, Gibbons BJ, Davis RE, Nagai S, Liu X, Robinson PJ, Wu T, Kaplan CD, Kornberg RD. Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proc Natl Acad Sci U S A. 2012; 109:4816–4821. [PubMed: 22411836]

Oksenych V, Bernardes de Jesus B, Zhovmer A, Egly JM, Coin F. Molecular insights into the recruitment of TFIIH to sites of DNA damage. EMBO J. 2009; 28:2971–2980. [PubMed: 19713942]

Okuda M, Tanaka A, Satoh M, Mizuta S, Takazawa M, Ohkuma Y, Nishimura Y. Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. EMBO J. 2008; 27:1161–1171. [PubMed: 18354501]

Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, Aebersold R. Identification of TFB5, a new component of general transcription and DNA repair factor TFIIH. Nat Genet. 2004; 36:707–713. [PubMed: 15220919]

Rossignol M, Kolb-Cheynel I, Egly JM. Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. EMBO J. 1997; 16:1628–1637. [PubMed: 9130708]

Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS biology. 2012; 10:e1001244. [PubMed: 22272186]

Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology. 1993; 234:779–815. [PubMed: 8254673]

Sandrock B, Egly JM. A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH. J Biol Chem. 2001; 276:35328–35333. [PubMed: 11445587]

Schmitt DR, Kuper J, Elias A, Kisker C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. PloS one. 2014; 9:e102389. [PubMed: 25013903]

Schultz P, Fribourg S, Poterszman A, Mallouh V, Moras D, Egly JM. Molecular structure of human TFIIH. Cell. 2000; 102:599–607. [PubMed: 11007478]

Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic acids research. 2005; 33:W244–W248. [PubMed: 15980461]

Springer TA. Complement and the multifaceted functions of VWA and integrin I domains. Structure. 2006; 14:1611–1616. [PubMed: 17098186]

Thomas MC, Chiang CM. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 2006; 41:105–178. [PubMed: 16858867]

Tirode F, Busso D, Coin F, Egly JM. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell. 1999; 3:87–95. [PubMed: 10024882]

Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends in biochemical sciences. 2012; 37:509–516. [PubMed: 22989858]

Wang F, Chang D, Hu FL, Sui H, Han B, Li DD, Zhao YS. DNA repair gene XPD polymorphisms and cancer risk: a meta-analysis based on 56 case-control studies. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008; 17:507–517.

Whittaker CA, Hynes RO. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. Molecular biology of the cell. 2002; 13:3369–3387. [PubMed: 12388743]

Yee A, Nichols MA, Wu L, Hall FL, Kobayashi R, Xiong Y. Molecular cloning of CDK7-associated human MAT1, a cyclin-dependent kinase-activating kinase (CAK) assembly factor. Cancer research. 1995; 55:6058–6062. [PubMed: 8521393]

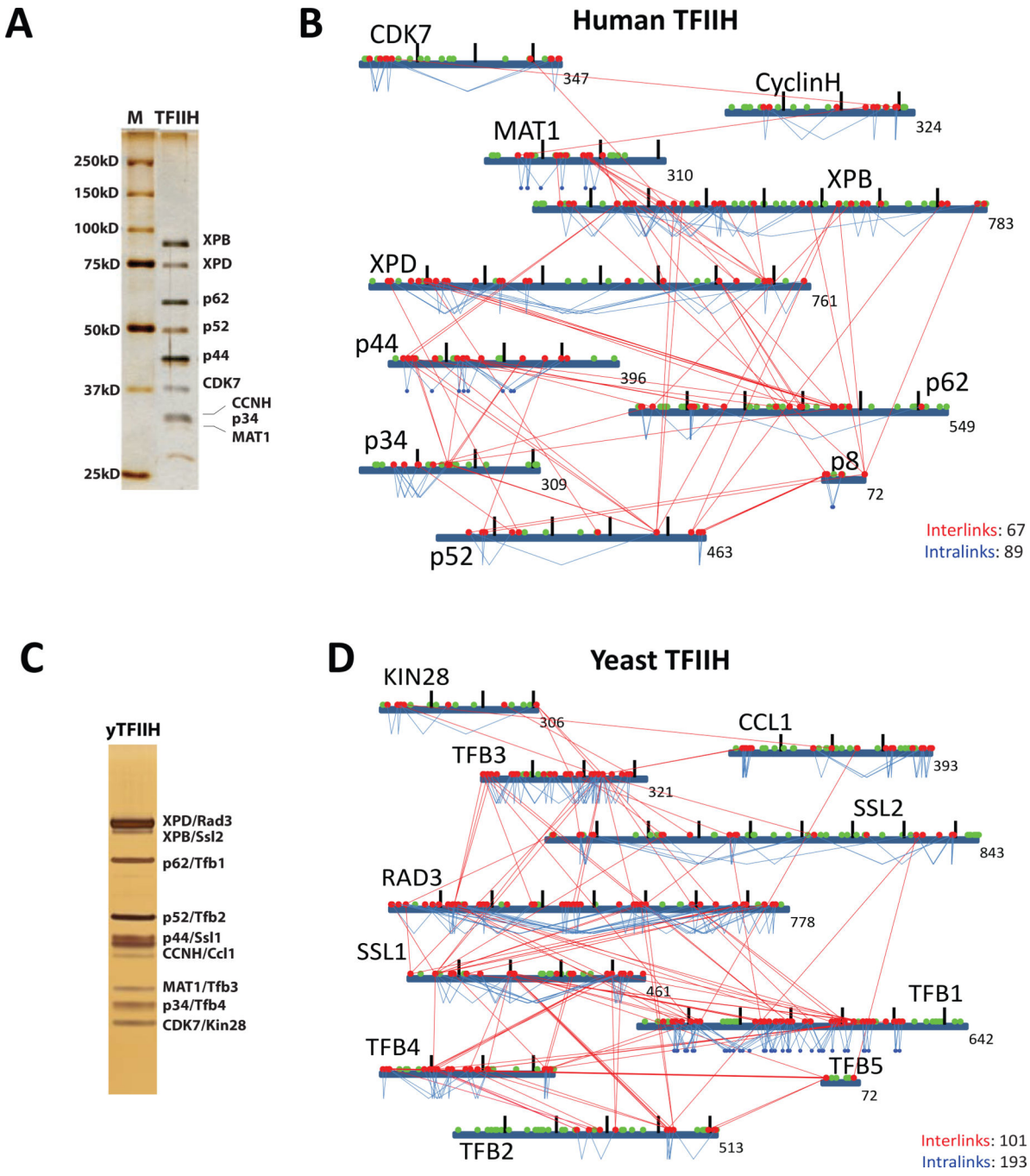

Figure 1. Purification and crosslinking maps of human and yeast TFIIH

A) Silver stained gel of purified human TFIIH. **B**) Map of the identified inter-protein (red lines) and intra-protein (blue lines) crosslinks for human TFIIH. **C**) Silver stained gel of purified yeast TFIIH. **D**) Inter- and intra-protein crosslink map for yeast TFIIH as in (**B**). Red and green dots indicate the positions of lysine residues. Red dots indicate that the lysine residue was identified in a crosslink.

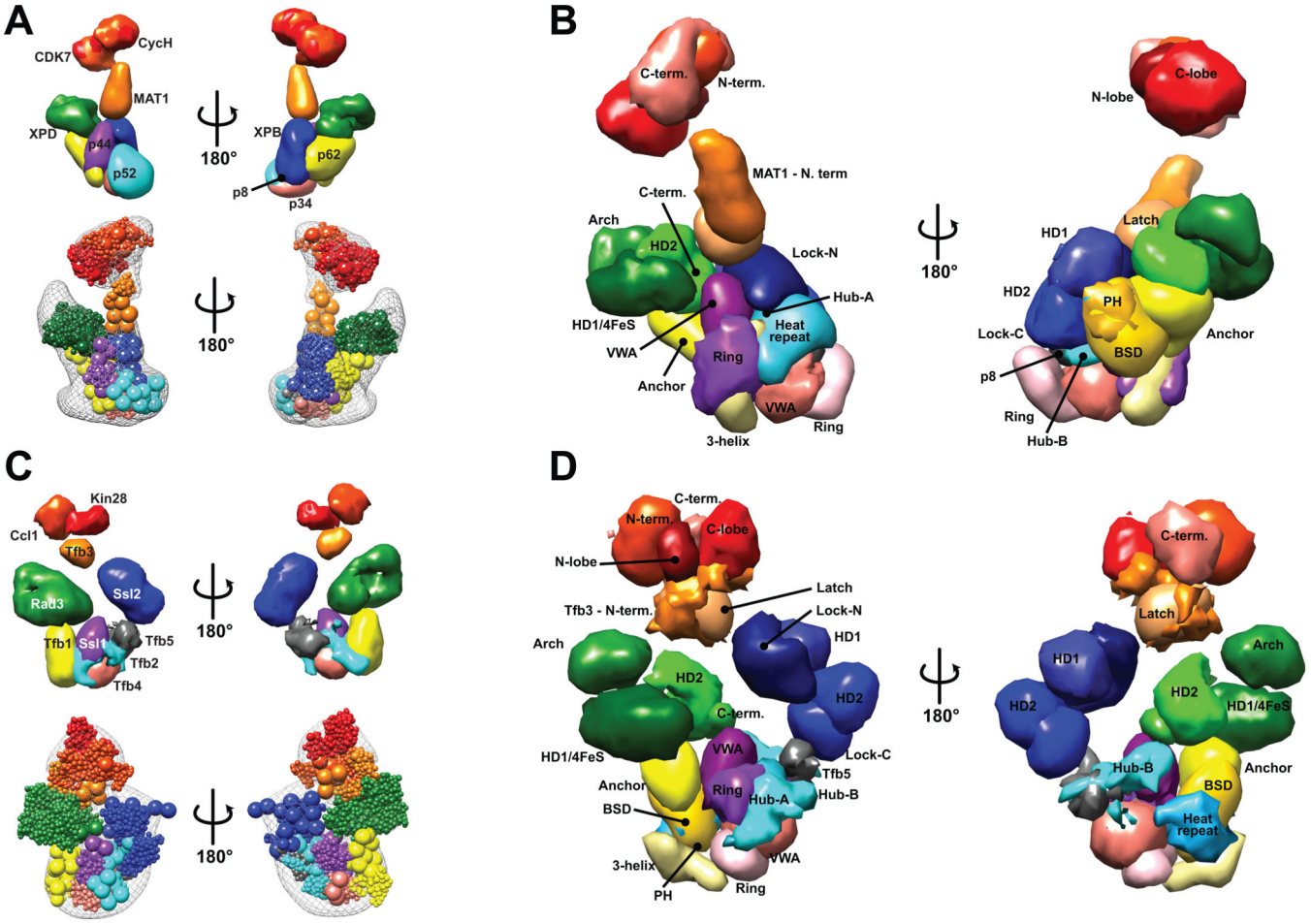

Figure 2. Molecular architecture of human and yeast TFIIH

A) Localization density map of the human core TFIIH subunits (top) and the best scoring model (bottom). The EM density map of human TFIIH used in this study is shown in grey mesh. **B**) Domain decomposition of human core TFIIH. **C**) Localization density map of the yeast TFIIH subunits (top) and the best scoring model (bottom). The EM density map of yeast TFIIH used in this study is shown in grey mesh. **D**) Domain decomposition of yeast TFIIH. The human and yeast models are superposed on the XPD/Rad3 subunit. The different shapes between the two models are due to differences in the cryo-EM density maps and may not represent actual structural differences between the two species, but are possibly due to artifacts due to different sample preparations and data collection as well as processing.

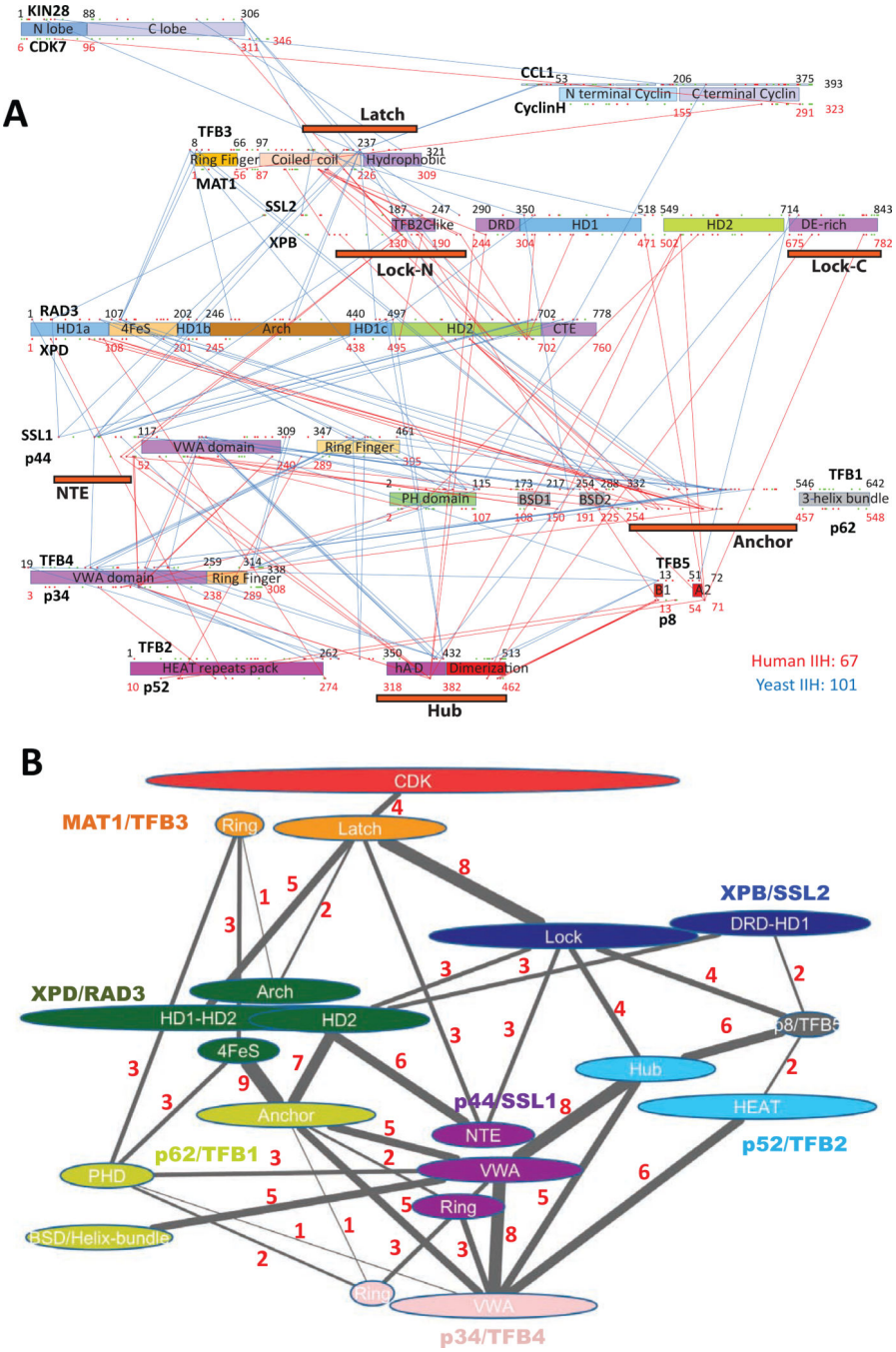

Figure 3. Summary of CXMS results for human and yeast TFIIH

A) Map of the identified inter-protein crosslinks for human (red lines) and yeast TFIIH (blue lines). Aligned sequences of human (bottom) and yeast (top) subunits are shown. Known and predicted structural domains are indicated as well as the positions of conserved topological regions identified in this study (orange bars). **B**) Summary of identified domain-domain crosslinks for human and yeast TFIIH. The thickness of the line is proportional to the number of crosslinks identified for each domain-domain linkage. The number of crosslinks supporting each domain-domain linkage is provided.

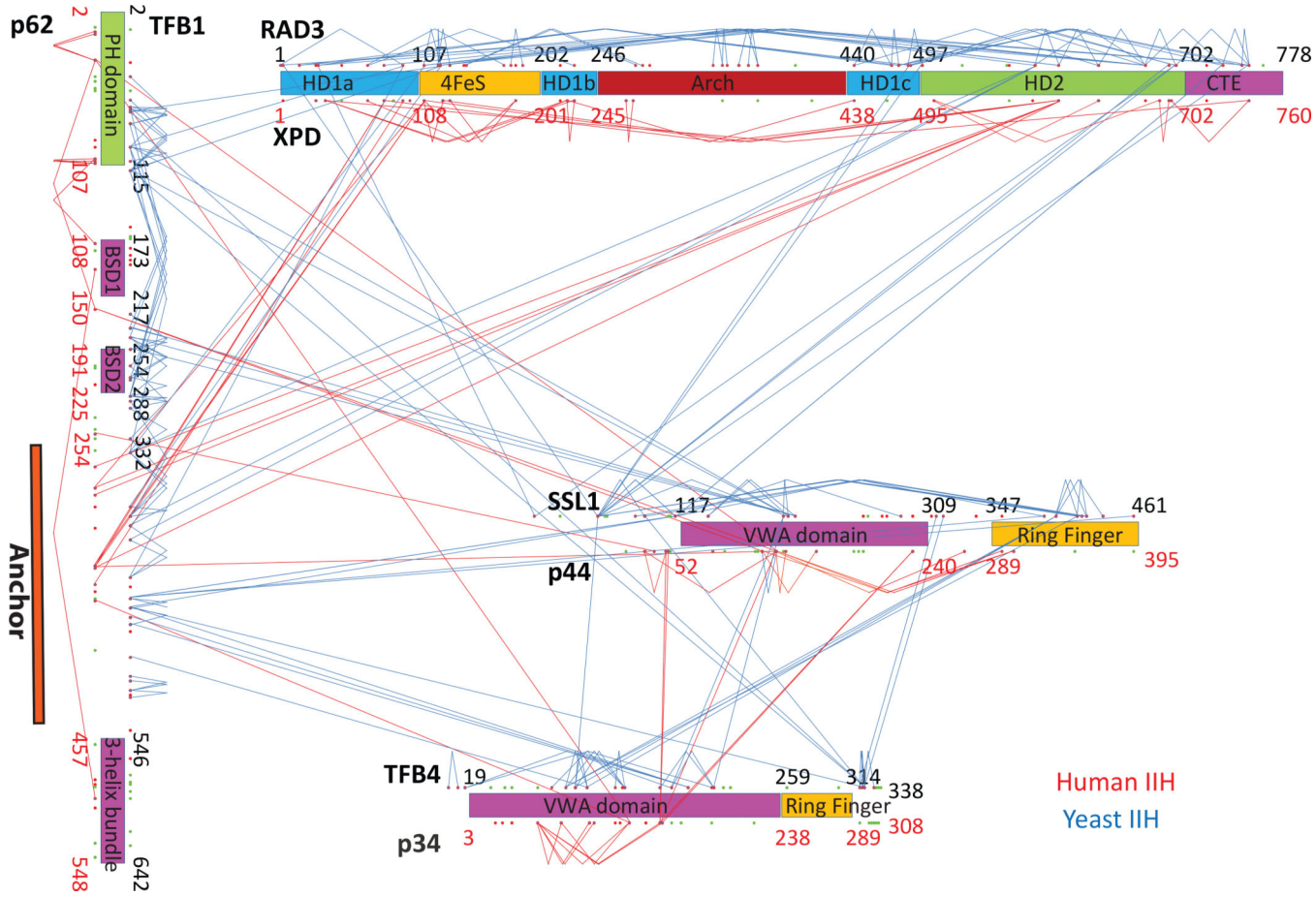

Figure 4. Crosslinks involving p62/Tfb1, XPD/Rad3, p44/Ssl1, and p34/Tfb4

Identified inter- and intra-protein crosslinks for human (red lines) and yeast (blue lines) homologs are mapped onto their aligned sequences. Known and predicted structural domains are indicated as well as the position of the p62/Tfb1Anchor region (red bar).

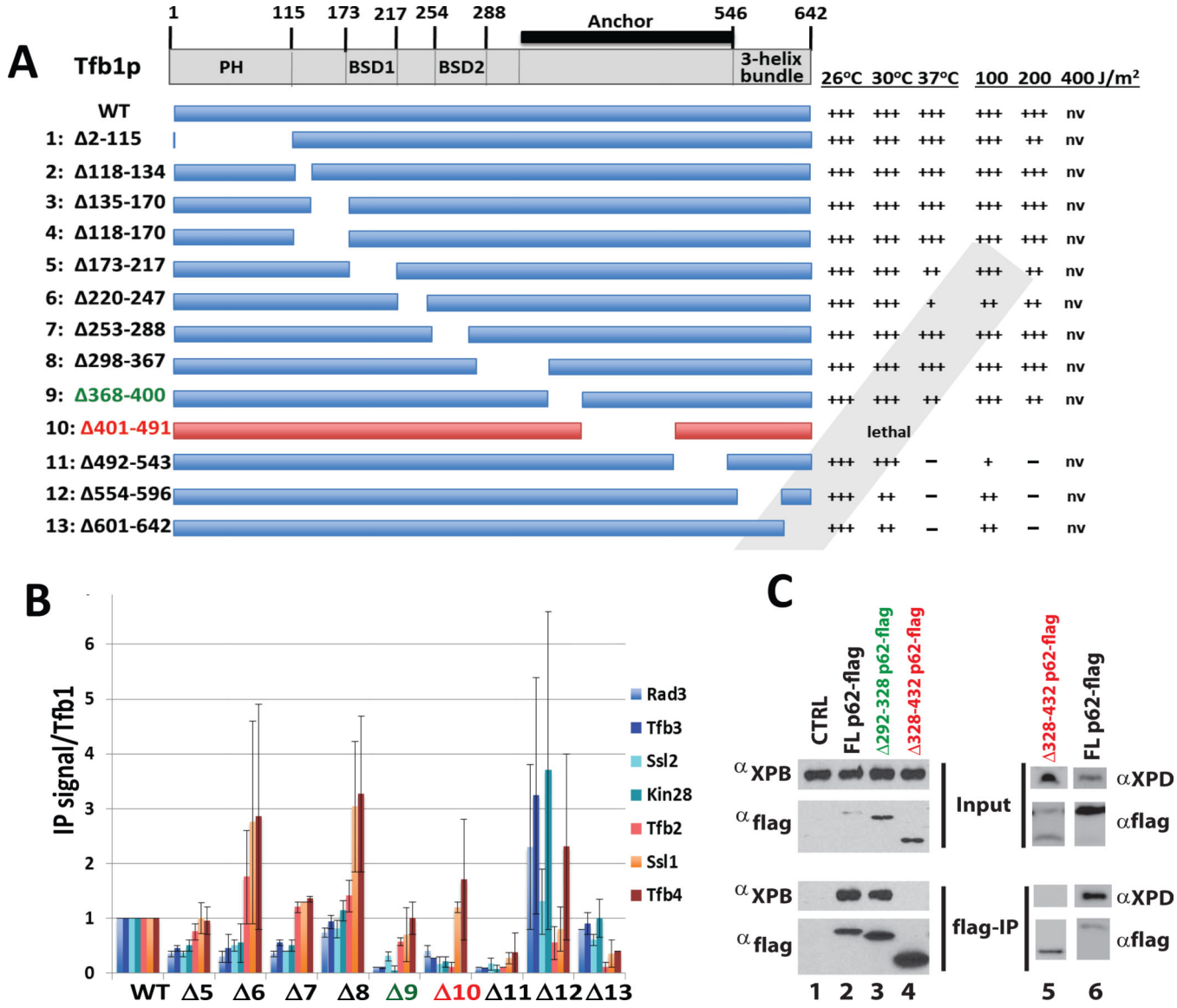

Figure 5. The p62/Tfb1 Anchor region is essential for the structural integrity of TFIIH

A) Schematic of the Tfb1 serial deletions analyzed in this study along with results of growth and UV sensitivity assays. Red bar indicates lethal phenotype. **B**) Results of IP-western analyses of Tfb1 domain deletions. The results are normalized to IP levels in the wild type Tfb1 strain. Error bars represent SEM for n = 3. **C**) Co-immunoprecipitation analysis with flag-tagged WT p62 (lane 2) or Anchor region deletions 292–328 (lane 3) or 328–432 (lane 4) in HeLa cells. CTRL: untransfected cells (lane 1). Note human p62 residues 292–328 and 328–432 align to yeast Tfb1 residues 359–397 and 397–505, respectively.

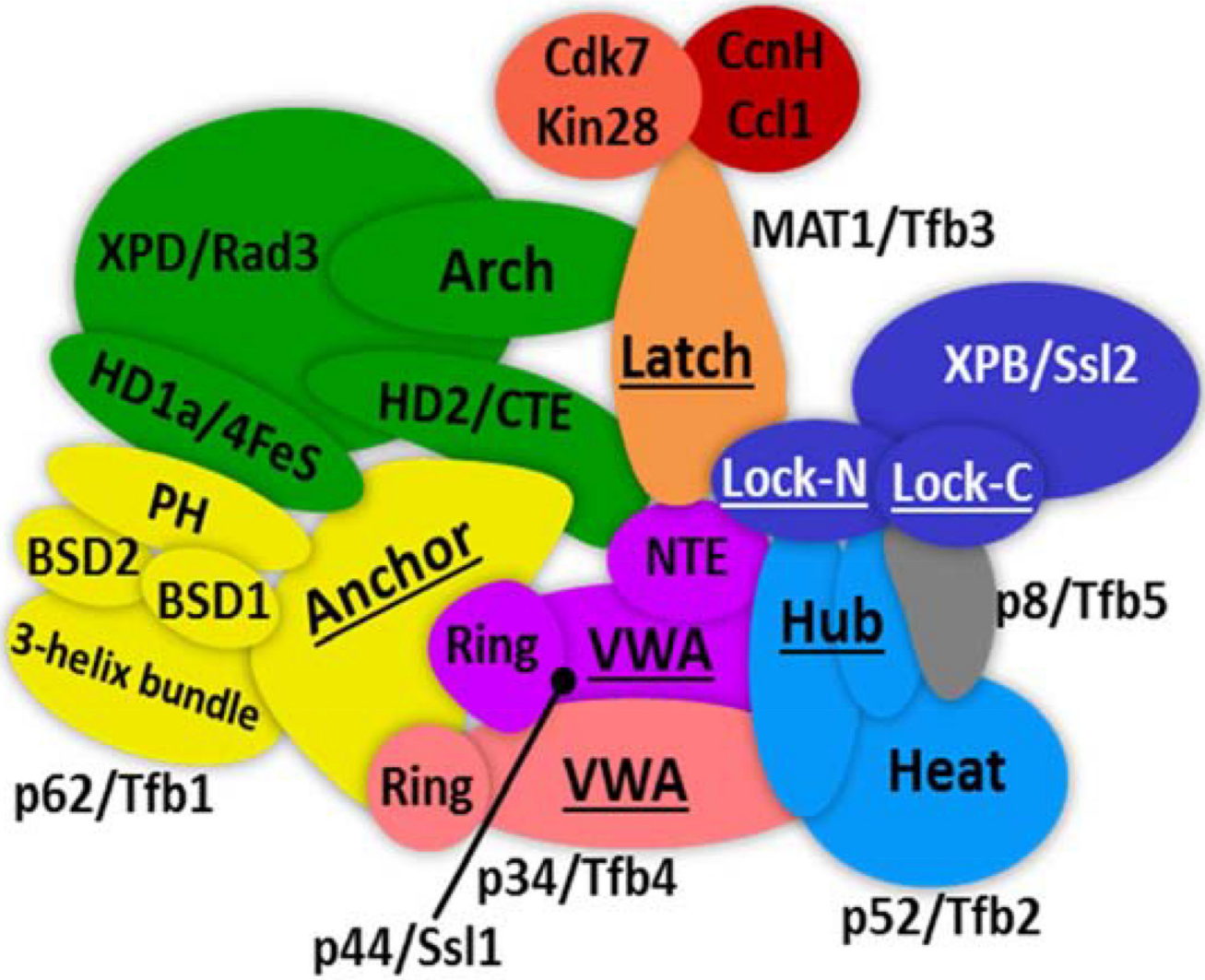

Figure 6. Schematic model of TFIIH subunit organization highlighting conserved topological features

The conserved topological features identified in this study are underlined. Subunit coloring scheme is the same as in Figure 2.

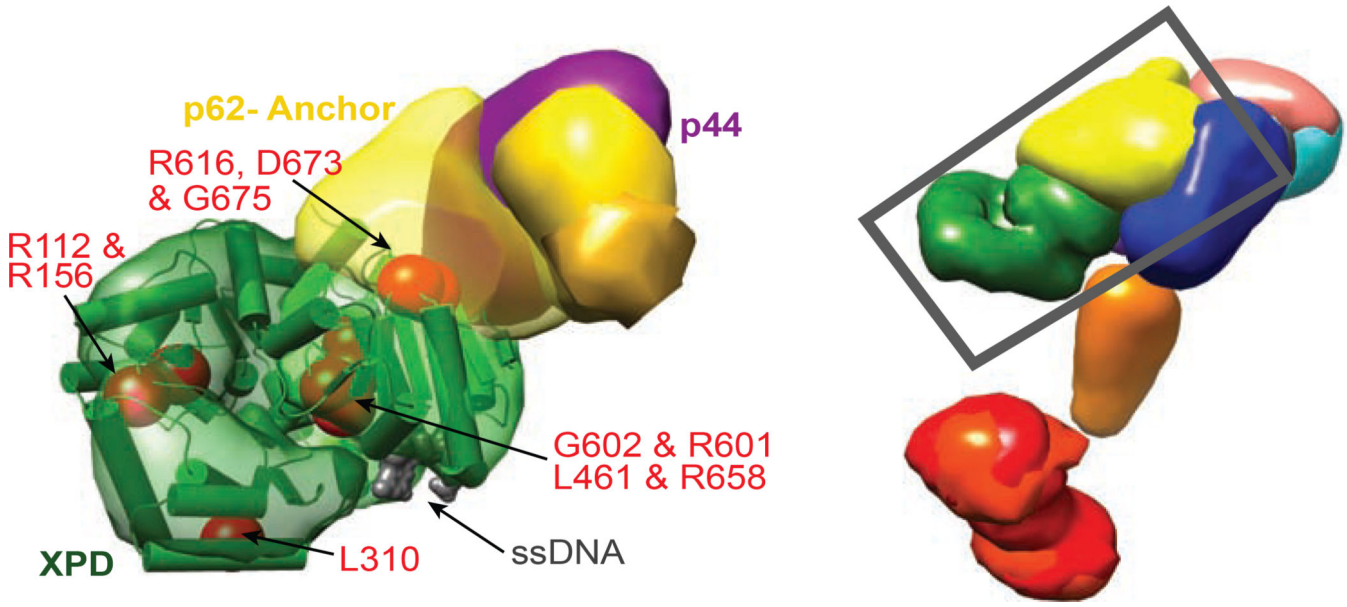

Figure 7. Mapping disease-associated mutations in XPD onto the human TFIIH models

Average placement of XPD without its C-terminal domain (precision of 8.5 Å) in the human TFIIH models is shown on the left. Disease-associated mutations in XPD are shown as red spheres. The location of the subunits in the context of core TFIIH is indicated by the grey rectangle on the right.