SUPPLEMENTARY INFORMATION

SUPPLEMENTARY NOTES

New positioning of the cyclin-activating kinase subcomplex is revealed in the holo-PIC structure

Another significant aspect of our model is that we were able to shed light on the positioning of the cyclin-activating kinase subcomplex CAK comprised of CDK7, Cyclin-H and the hydrophobic C-terminal end of MAT1. Figure S2 and Movie S2 highlight the location of CAK in the EM density of the closed-complex human PIC. Within the CAK module CDK7 associates with Cyclin-H whose primary role is to regulate kinase activity and direct CDK7 in its search for substrates. The role of MAT1 in the assembly appears to be purely structural: the subunit serves to tether the CAK module to the TFIIH core and to further stabilize the interaction of the CDK7/Cyclin-H pair. Indeed, our model suggests that upon exiting the helical bundle (residues 126-213) the MAT1 protein chain is directed toward the CAK module via an extended linker. The C-terminal end of MAT1 was predicted to contain only two structured helical segments, whose positions we were able to identify in the EM density. The rest of the sequence is characterized by unstructured linkers allowing the entire MAT1 C-terminus, including the two helices, to wrap around the CDK/Cyclin pair. The CAK subcomplex is multifunctional and manifestly important not only for transcription regulation but more broadly for the regulation of RNA processing in the cell. Specifically, the CAK module phosphorylates the regulatory unstructured C-terminal domain (CTD) of Rpb1 in Pol II. The CTD contains 52 heptad repeats. Phosphorylated CTD then functions as a scaffold in the recruitment of proteins involved in such diverse cellular activities as transcriptional pausing, mRNA capping, splicing, processing of spliceosomal small nuclear RNAs and RNA polyadenylation. Surprisingly, in our model CAK is positioned next to Rpb4 and in close proximity to the C-terminal end of Rpb1. Indeed, in the EMD-3307 reconstruction we identified a region of weak density between the kinase module and Rpb1 (Figure S2c and S2d) that cannot be attributed to any structured PIC component and is likely the globular unstructured CTD domain of Rpb1. Thus, CAK appears to be in an ideal position to phosphorylate its CTD molecular target. Finally, in Figure S2e, S2f and S2g we compare the CAK module location in apo-TFIIH, holo-PIC (our model) and the yeast PIC/TFIIH/DNA complex. The CAK position in all three cases appears to be highly variable. In apo-TFIIH CAK is proximal to the MAT1 ARCH anchor domain. However, EM has shown that this position is flexible, allowing CAK to slide up and down the long MAT1 helix. By contrast, in the yeast complex CAK was found to occupy a position adjacent to the large co-activator Mediator and is completely separated from the rest of the PIC. The model advanced

by Schilbach et al. suggest that the Pol II CTD may extend along a groove formed by Mediator and Pol II toward the CAK and that phosphorylation by Kin28 (yeast homolog of CDK7) could, thus, triggers Mediator dissociation from the PIC. This would allow the elongation complex to progress past the TSS. Strikingly, all three EM reconstructions show vastly different CAK subcomplex positions, indicating that CAK may shuttle among various components of the transcription machinery and that its mode of action is likely context-dependent.

SUPPLEMENTARY TABLES

Table S1 Summary of TFIIH and PIC structural elements and original sources used for hybrid modeling.

Protein	Chain	Size (aa)	Modeled Residues	Alternative names	Structures (PDB IDs) used for hybrid modeling
XPD	0	760	11-742	ERCC2	Residues 11-742 modeled from 5OF4
p62	1	548	1-546	GTF2H1	Residues 159-546 built de novo; p62 BSD1 domain (residues 110-158) modeled from NMR structure (2DII); p62 PHD domain (residues 1-109) modeled from NMR structure (2RNR)
p52	2	462	6-458	GTF2H4	Residues 284-384 built de novo; Residues 6-383 constructed by homology modeling using the yeast Tfb2 (5OQJ) as a template
MAT1	3	309	1-309	MAAT1	Residues 65-309 built de novo; MAT1 RING domain (residues 1-64) modeled from NMR structure (1G25)
p34	4	308	6-300	GTF2H3	Residues 233-251 built de novo; ZINC finger domain (residues 252-300) constructed by homology modeling using the yeast Tfb2 (5OQJ) as a template; vWA domain (residues 6-251) modeled from 5OF4
p8	5	71	2-67	GTF2H5	Residues 2-67 modeled from 5OF4
p44	6	395	10-394	GTF2H2	Residues 1-57, 313-343 built de novo; vWA domain (residues 58-312) modeled from 5OF5; ZINC finger domain (residues 344-394) constructed by homology modeling using the yeast Ssl1 (5OQJ) as a template
XPB	7	782	30-201 267-728	ERCC3	Residues 30-201 and 267-300 built de novo; Residues 301-728 modeled from 5OF5
CDK7	8	346	13-311	MO15	Residues 13-311 constructed by homology modelling using human CDK2 (1JSU) as a template
CyclinH	9	323	11-286	CCNH	Residues 11-286 modeled from human Cyclin H (1KXU)
TFIIEα	Q	439	10-439	GTF2E1	Residues 10-215 modeled from human TFIIE (5GPY); Residues 215-234 built de novo; Residues 335-439 modeled from NMR structure 2RNR
TFIIEβ	R	292	75-242	GTF2E2	Residues 75-242 modeled from NMR structure (2RNR)
Core PIC (pol II, TBP, TFIIA, TFIIB, TFIIF, TFIIS and DNA)					Core PIC modeled from the EM structure (5IY6)

Table S2 Summary of interfaces between TFIIH and core-PIC.

Interface	Area (Å^2) [1]	N_{inter} [2]	N_{HB} [3]	N_{SB} [4]
MAT1 – Rpb4/7	584.3	17	11	9
MAT1 – TFIIEα	320.4	13	5	5
CDK7 – Rpb4	456.7	15	3	2
TFIIEα – p62	4607.2	105	75	41
TFIIEα – TFIIH	6107.6	141	97	56
TFIIEα – core-PIC	1257.0	43	16	8
TFIIEβ – core-PIC	628.7	20	9	6

1 – Buried surface area at the interface
2 – Number of interfacial residues
3 – Number of interfacial hydrogen bonds
4 – Number of interfacial salt bridges

Table S3 Human disease mutations in TFIIH – Model and Community-based analysis.

Protein	Mutation	Disease*	C ID‡	Inferred Function†	Reference, Activity, Structural Analysis, Inferred Change
P8	L21P	TTD	P	Protein-protein XPB interaction	DOI: 10.1038/ng1387. L21P is active for transcription but NER defective, DOI: 10.1038/ng1387. Loses structure DOI: 10.1038/nsmb.1478. L21 acts in hydrophobic and helix-helix packing against XPB. Pro mutations likely cause defects in the protein interface and fold.
XPB	F99S	XP/CS	O	Community connector	PMID:8304337. F99 is in a helix that stacks on T119 (TTD mutation). Nearby beta bulge Pro impacts interaction with p52 and p44. Mutation to serine likely destabilizing to community.
XPB	T119P	TTD	O	XPB interaction to p52 and p44	Weeda, 1997 (PMID: 9012405). T is at the end of a beta strand and stacks on F99 (XP/CS mutation). T to P is destabilizing. See F99S. This XPB mini-domain interacts with p52 and p44.
XPD	Y18H	XP/CS, TTD	A	Community connector	DOI: 10.4161/rdis.24932. Y18 is at the end of a helix and H-bonds to G45 backbone. (Two residues from XP/CS mutation at position 47). H mutation between two helicase domains likely breaks E20 H-bonding network. The phenotype resulted from compound heterozygote mutations with Y18H likely associated only with the XP/CS features. The second mutation was A725P, a known TTD mutation.
XPD	G47R	XP/CS, COFS††	A	Community connector	DOI: 10.1111/j.0022-202X.2005.23745.x, 10.1093/hmg/3.10.1783. G47 in a helicase motif ends helix in DNA binding groove near Y18. R mutation likely impacts helix and carboxylates in p62 interface.
XPD	S51F	XP/TTD	A	ATPase function, Protein-protein p62 interaction	DOI: 10.1002/humu.20768. S51 packs in tight pocket with H-bonds to backbone carbonyl (res 48 i-3 – next to G47 XP/CS mutant side) and caps the helix. See DOI: 10.1016/S0006-3495(00)76514-3. Mutation to F will distort the region and helical end near interface with p62, consistent with a TTD phenotype. S51

					neighbors the ATPase motifs, and mutation to F is likely to disrupt ATPase activity.
XPD	T76A	XP	A	ATPase Function	DOI: 10.1101/gad.859501. T76 helps end helix and stabilizes metal binding carboxylates in Walker B. Alanine mutation cuts helicase activity (DOI: 10.1016/j.cell.2008.04.030) by destabilization.
XPD	R112H/C	XP/TTD	A	Indirect DNA binding, Protein-protein p62 interaction	DOI: 10.1101/gad.859501, 10.1002/humu.20768, 10.1093/hmg/10.22.2539. R112 spans helices, protects the Fe-S cluster, stabilizes the interface with the XPD arch domain and the DNA binding groove. Mutation reduces helicase activity (DOI: 10.1016/j.cell.2008.04.030) and thus NER. R112 is close to the interface, consistent with an assembly defect leading to TTD
XPD	D234N	XP	A	ATPase function	DOI: 10.1101/gad.859501. D234 (in the Walker B motif) binds Mg for ATP hydrolysis. N mutation cuts helicase activity (DOI: 10.1016/j.cell.2008.04.030) for repair.
XPD	C259Y	TTD	D	Mat1 interaction	DOI: 10.1101/gad.859501. C259 caps a helix in hydrophobic pocket. Mutation to bulky Y residue destabilizes region proximal to Mat1 interface.
XPD	R378H	TTD	D	Mat 1 interaction	DOI: 10.1002/humu.20768. R378 is in helix packing with L382 and H328 with amides H-bonding to S322 and A324 backbone carbonyl. Mutation to H destabilizes helix position for Mat1 interaction.
XPD	L461V	XP/CS	A	Community disruptor	DOI: 10.1038/jhg.2015.18, 10.1111/j.0022-202X.2005.23745.x, 10.1016/j.ccr.2006.05.027. L461 in helicase motif packs against F654 and L456. Val mutation impacts helicase and makes region more rigid due to branched beta carbon.
XPD	L485P	XP	A	Helicase domain	DOI: 10.1093/hmg/10.22.2539. L485 is in a loop, packing with A732 near the helicase domain interface. Pro mutation alters backbone geometry, disrupting XPD C-terminus near R487, a TTD mutant.
XPD	R487G	TTD	A	p44 protein interaction	DOI: 10.1101/gad.859501. R487 in a loop near p44 interface provides aliphatic stacking and H-bonding. Mutation to G removes side chain interactions and main chain stability.
XPD	R511Q	XP	A	DNA binding	DOI: 10.1101/gad.859501. R511 points into the DNA binding groove DNA. Mutation to Q likely to reduce DNA binding.
XPD	S541R	XP	A	DNA binding groove	DOI: 10.1101/gad.859501. S541 next to Y542C (XP) is in a helicase motif. It caps a helix flanking the DNA binding groove. Mutation to R likely reduces helix cap stability and impacts DNA binding near XPB and p62 interfaces.
XPD	Y542C	XP	A	DNA binding	DOI: 10.1101/gad.859501. Y542 in a helicase motif packs against aliphatic chains of Q572 and V599 to orientate a helix N-terminus. Mutation to C creates a destabilizing void under the DNA-binding groove. It is near S541 (XP) and XPB and p62 binding interfaces.

XPD	R592P	TTD	A	Conformation, p44 and p62 interface	DOI: 10.1101/gad.859501. R592 in a helicase motif is a probable C-cap for a helix and in the interface with p62. R592 and A594 are in a loop at the junction of XPD with p44 and p62. Mutation to P removes interactions and likely disrupts this region by backbone distortion.
XPD	A594P	TTD	A	p44 and p62 interaction	DOI: 10.1101/gad.859501. A594 is in a helicase motif at the junction of XPD, p44 and p62 interfaces and packed with F568 and C588 in a hydrophobic pocket. Mutation to P would disrupt packing and backbone conformation at the protein interfaces.
XPD	R601L/W	XP	A	Helicase movement	DOI: 10.1101/gad.859501. R601 is in a helicase motif loop near two XP mutations (541 and 542) and the DNA binding groove neighboring p62 and XPB interfaces. R aliphatic chain packs against H659 and V626 near the interface between helical domains and the Walker B motif. L or W mutation likely disrupts packing and helicase activity (see DOI: 10.1016/j.cell.2008.04.030).
XPD	G602D	XP/CS	A	Dynamic community disruptor	DOI: 10.1101/gad.859501. G602 is in helicase motif loop at end of a helical turn adjacent to the DNA binding groove and near p62 and XPB interfaces. D mutation will reduce structural flexibility at the DNA and protein interfaces.
XPD	R616P/Q	TTD	A	Protein-protein P44 interaction	DOI: 10.1111/exd.12166. R616 in a loop at the end of a beta strand H-bonds to p44 E204. Its chain stacks against Y674 and P532. Mutation disrupts the salt bridge and the p44 interface.
XPD	R636W	TTD	D	Protein-protein XPB interaction	Shin, 2013, PMID: 23884229. R636 is in a helix at the interface with a helical arm of XPB. Its aliphatic chain packs against L640 and M438. Its guanidinium H-bonds to D240 and N241 near XP mutant R658. W mutation would sterically clash causing helix and XPB interface distortion
XPD	R658H/C	TTD	A	Protein-protein XPB interaction	DOI: 10.1101/gad.859501. Takayama, 1996, 8571952. R658 is in a helicase motif and helix at the interface of XPD helicase domains and an XPB interface. It makes a salt bridge to D240 next to the XPD Walker B motif. Its side chain packs against F654 and F651 adjacent to the XPB interface. H or C mutation will prevent the packing to the two F side chains and weaken the XPB interface.
XPD	C663R	XP/TTD	A	ATPase function, Protein-protein p62 interaction	DOI: 10.1101/gad.859501, 10.1002/humu.20768. C663 is in a helicase motif and helix with its side chain in a hydrophobic pocket and packing with F538 and A600. R mutation likely distorts the helix and p62 interface.

XPD	R666W	XP/CS	A	Community flexibility disruptor	DOI: 10.1101/gad.859501, 10.1038/jhg.2015.18, 10.1111/j.0022-202X.2005.23745.x, 10.1016/j.ccr.2006.05.027. R666 is in a helicase motif at the end of a helix. It extends into a pocket where it forms a salt bridge to E235 in Walker B motif. W mutation removes ATPase and helicase activity (see DOI: 10.1016/j.cell.2008.04.030) and disrupts interactions while increases local rigidity by filling the pocket.
XPD	D673G	TTD	A	Protein-protein p62 (and p44)	DOI: 10.1101/gad.859501. D673 on a loop at the end of a beta strand with likely interactions to R669, H613, and/or G670. R669 forms the interface with p62 near p44 contacts. G mutation would disrupt the H-bond network and p62 interaction.
XPD	G675R	XP/CS	A	Community disruptor	DOI: 10.1101/gad.859501. G675 lies at the end of a beta strand. R mutation would fill an adjacent pocket reducing helicase activity (DOI: 10.1016/j.cell.2008.04.030) and increasing rigidity.
XPD	D681N/H	XP	A	Helicase coordination, interface to XPB	DOI: 10.1101/gad.859501, 10.1002/humu.20768. D681 at the end of a beta strand H-bonds to R683(XP). Both are located on a loop between a strand and helix. R683 contacts XPB. N or H mutation likely disrupts H-bonding and coordination of helicase activities for NER.
XPD	R683W/Q	XP	D	Protein-protein XPB interface, helicase coordination, ssDNA binding	DOI: 10.1101/gad.859501, 10.1002/humu.20768, 10.1073/pnas.94.16.8658. R683 lies on a loop between a strand and helix and contacts XPB. R to W mutation reduces ATPases, helicase and DNA binding ((DOI: 10.1016/j.cell.2008.04.030). W or Q mutation likely disrupts interaction interface and coordination of helicase activities for NER explaining disease phenotypes. Intriguingly, overlay with a recent DinG structure reveals the ssDNA binding surface of XPD family proteins, and R683 maps to this surface. The XPB helix blocks this ssDNA binding path. Mutation would likely have an effect on ssDNA binding and potentially XPB regulation of XPD.
XPD	G713R	TTD	A	Protein-protein XPB interaction	DOI: 10.1101/gad.859501. G713 is in the middle of a helix near a hydrophobic core. R mutation would sterically push into the core disrupting the helix at the XPB interface.
XPD	A717G	XP	A	Helicase coordination	DOI: 10.1038/jhg.2015.18. A717 is right below G713 in the same helix and its methyl goes into the same hydrophobic core formed by V530, V714, L721, L676, M493, and M527. G mutation next to G713 and R722 (XP) likely makes the helix overly flexible uncoupling signaling with XPB.
XPD	R722W	XP/TTD, TTD	A	Helicase coordination via p44 and p62	DOI: 10.1101/gad.859501, 10.1002/humu.20768. R722 on the same helix with A717(XP) makes a salt bridge to D75 of p44, which may facilitate NER as this helix is also adjacent the p62 cap on the XPD ATP binding site. W mutation would disrupt the p44

					interface near p62. N. B. P44 and XPD in this region are in different communities.
XPD	A725P	TTD or XP/CS, TTD	A	Community disruptor, reduce flexibility	DOI: 10.1101/gad.859501, 10.4161/rdis.24932. A725 is a C-cap of a helix at the p44 interface. P mutation makes a good N-cap that should disrupt helix stability but increase rigidity. P44 and XPD in this region are in different communities. Reference 10.4161/rdis.24932 classifies the mutation as TTD only. Reference 10.4161/rdis.24932 describes a complex phenotype due to two mutations Y18H (likely associated XP/CS features) and A725P, associated with TTD.
TFIIEβ	A150P	TTD	E	Protein-protein interaction between TFIIEα and β subunits	DOI: 10.1016/j.ajhg.2016.02.008. A150 is in the WH2 domain of TFIIEβ is at the N-terminal end of a helix. Mutation to proline likely disrupts the helix and particularly the packing of a neighboring K148 with L185 and A181 on an adjacent helix. The mutation will likely disrupt the assembly of TFIIE subunits.
TFIIEβ	D187Y	TTD	E	Protein-protein interaction between TFIIEα and β subunits	DOI: 10.1016/j.ajhg.2016.02.008. D187 is in the WH2 domain of the TFIIEβ subunit is located at the carboxy terminus of an alpha helix. It forms a salt bridge with K206, located in a coil region that packs against TFIIE alpha subunit. Mutation to Y disrupts the salt bridge and the positioning of the coil region with the alpha subunit, disrupting assembly of TFIIE subunits.

* Disease abbreviations: Xeroderma pigmentosum (XP), trichothiodystrophy (TTD), Xeroderma pigmentosum/Cockayne syndrome (XP/CS)

‡ C ID denotes dynamic community identifier. Communities are defined from dynamic network analysis as outlined under Methods.

† Inferred function is the functional role of the mutation inferred from our structural and dynamic model.

††COFS – denotes cerebro-oculo-facio-skeletal syndrome

Table S4 MolProbity results of apo TFIIH and holo PIC.

Validation	Apo TFIIH	Holo PIC
MolProbity score	2.66	2.12
MolProbity Clashscore	23.5	7.98
Rotamers outliers (%)	1.34	0.39
Cβ deviations (%)	0.23	0.04
Ramachandran favored (%)	81.34	83.72
Ramachandran allowed (%)	17.17	13.33
Ramachandran outliers (%)	1.50	2.95

Table S5 Human and yeast TFIIH MDFF map-to-model statistics.

	MolProbity	iFSC1 (Å)[2]	iFSC2 (Å)[2]	d_Model (Å)	d_FSC_Model (Å)[3]
hTFIIH – h1	2.66	3.76	3.68	6.7	3.6/4.4/7.3
hTFIIH – h2	2.66	3.83	3.76	6.8	3.7/4.4/7.3
hTFIIH – f	2.66	4.50	4.29	4.2	3.9/4.3/7.2
yTFIIH – h1	2.46	2.82	2.80	6.7	3.9/6.4/8.7
yTFIIH – h2	2.46	2.77	2.75	6.8	4.0/6.2/8.6
yTFIIH – f	2.46	3.33	3.29	6.8	4.7/5.0/8.5

1 – All values are reported for both half maps and full map denoted h1 (half map 1), h2 (half map 2) and f (full map). hTFIIH denotes human TFIIH and yTFIIH denotes yeast TFIIH maps, respectively.
2 – Integrated FSCs (iFSC) between 12 – 4.4 Å for human TFIIH (hTFIIH) and 12 – 4.7 Å for yeast TFIIH (yTFIIH).
3 – Values reported at 0, 0.143 and 0.5 Fourier shell coefficients (FSC).