Summary of integrative structure determination of Structure of the pre-incision complex in nucleotide excision repair with tilted XPG core (PDB ID: 9A89 | pdb_00009a89, PDB-Dev ID: PDBDEV_00000374)

1. Model Composition	
1.1. Entry composition	- General transcription and DNA repair factor IIH helicase subunit XPB: chain(s) A (720 residues) - water: chain(s) AA [J] - General transcription and DNA repair factor IIH helicase subunit XPD: chain(s) B (760 residues) - General transcription factor IIH subunit 4: chain(s) C (441 residues) - General transcription factor IIH subunit 2: chain(s) D (377 residues) - General transcription factor IIH subunit 3: chain(s) E (292 residues) - General transcription factor IIH subunit 5: chain(s) F (66 residues) - DNA repair protein complementing XP-A cells: chain(s) G (273 residues) - General transcription factor IIH subunit 1: chain(s) H (154 residues) - DNA excision repair protein ERCC-5: chain(s) I (985 residues) - DNA repair endonuclease XPF Gene: ERCC4, ERCC11, XPF: chain(s) J (227 residues) - DNA excision repair: chain(s) K (198 residues) - Replication protein A 70 kDa DNA-binding subunit, N-terminally processed: chain(s) L (434 residues) - Replication protein A 14 kDa subunit: chain(s) M (115 residues) - Replication protein A 32 kDa subunit: chain(s) N (225 residues) - DNA (66-MER): chain(s) O [X] (66 residues) - DNA (66-MER): chain(s) P [Y] (66 residues) - IRON/SULFUR CLUSTER: chain(s) Q [B] - ZINC ION: chain(s) R [D], S [D], T [D], U [E], V [E], W [G], Y [L] - MAGNESIUM ION: chain(s) X [J], Z [Y]

1.2. Datasets used for modeling	- Experimental model, PDB: pdb_00006ro4 - 3DEM volume, EMDB: EMD-4970 - Experimental model, PDB: pdb_00006tuw - Crosslinking-MS data, Not available: 10.1038/s41467-019-10745-5 - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - Experimental model, PDB: pdb_00006sxa - Experimental model, PDB: pdb_00006sxb - Experimental model, PDB: pdb_00002bgw - De Novo model, Not available - Experimental model, PDB: pdb_00004gop - Experimental model, PDB: pdb_00006i52 - Experimental model, PDB: pdb_00001jmc - Experimental model, PDB: pdb_0000l1lo - Experimental model, PDB: pdb_00002jnw - Experimental model, PDB: pdb_00004mqv - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - De Novo model, Not available - De Novo model, MODEL ARCHIVE: ma-2chon
2. Representation	
2.1. Number of representations	1
2.2. Scale	Atomic
2.3. Number of rigid and flexible segments	0, 18
3. Restraints	
3.1. Physical principles	Information about physical principles was not provided
3.2. Experimental data	- 1 unique CrossLinkRestraint: BS3, 21 crosslinks - 1 unique EM3DRestraint: Rigid-body fitting using Chimera
4. Validation	
4.2. Number of ensembles	0
4.3. Number of models in ensembles	Not applicable
4.4. Number of deposited models	1
4.5. Model precision	Not available
4.6. Data quality	EMD-4970: resolution is 3.50 Å
4.7. Model quality: assessment of atomic segments	- Clashscore: 23.69 - Ramachandran outliers: 56 - Sidechain outliers: 25

4.8. Fit to data used for modeling	- Satisfaction of crosslinks: 28.57% - 3DEM q-score(s): 0.23 - 3DEM atom inclusion score(s): 0.37
4.9. Fit to data used for validation	Fit of model to information not used to compute it has not been determined
5. Methodology and Software	
1. *5.1. Method name*	Not available
5.3. Method description	To construct a model of the pre-incision complex (PlnC), we systematically examined the cryo-EM structures and densities of human apo-TFIIH, TFIIH/XPA/DNA, and XPF/ERCC1, the NMR structure of XPA-ERCC1, and the X-ray structures of the XPG catalytic core and RPA-ssDNA (RPA70, RPA32, and RPA14). The TFIIH/XPA/DNA structure (PDB ID: 6RO4 and EMDB accession code: EMD-4970) was the starting point for model building. The PInC hybrid model has an NER bubble size of 23 nucleotides, matching the 27-nucleotide optimal length of the excision products and the XPF and XPG incision patterns. FEN1 shares 30% sequence identity with the XPG catalytic core (PDB ID: 6TUR, 6TUW, and 6VBH). Thus, we modeled DNA-bound XPG based on the human FEN1/DNA X-ray structure (PDB ID: 5UM9). XPG positioning into the hybrid model was based on existing XL-MS data. In addition, positioning of the XPG core required placement of the 3' DNA junction 8 nucleotides away from the expected position of the DNA lesion near XPD's His135 residue. The two XPG gateway helices (GH1, residues 82-126) and (GH2, residues 734-761) were predicted with AlphaFold2 and positioned in the gap between XPD's Arch and Fe-S domains in accordance with the crosslink data. The XPD-anchor domain (residues 157-296) was predicted by AlphaFold2 and fitted into the TFIIH/XPA/DNA cryo-EM density. The loop connecting GH1 and the XPD-anchor was built with Modeller. To model XPF/ERCC1, we used the cryo-EM structures of XPF/ERCC1 (PDB ID: 6SXA and 6SXB). We first docked the XPF nuclease domain to the 5' junction. The catalytic metal was oriented 3A away from the scissile phosphodiester bond. Mg^{2+} ion coordination was based on the Aeropyrum pernix SNF2 structure (PDB ID: 2BGW). A water molecule was placed between Mg^{2+} ion and the DNA backbone phosphate group. The ERCC1 (HhH)2 domain was oriented to interact with the ssDNA through two DNA hairpins based on the 6SXB structure. The long linkers from the ERCC1 central domain to the ERCC1 (HhH)2 (residues 214-230) and from the XPF nuclease domain to the XPF (HhH)2 (residues 817-847) were built with Modeller. The SF2 helicase-like N-terminal domain of XPF was omitted from the hybrid PInC model due to lack of sufficient structural or biochemical restraints. To model RPA, we used following X-ray structures: Ustilago maydis RPA/ssDNA (PDB ID: 4GOP), yeast RPA/ssDNA (PDB ID: 6I52) and human RPA (PDB ID: 1JMC and 1L1O). The RPA70AB/ssDNA complex was modeled by superimposing the yeast RPA/ssDNA structure (PDB ID: 1JMC) onto the human apo-RPA

70AB (PDB ID: 6I52). Within PInC, only RPA70A, 70B, and 70C can engage DNA due to the size of the NER bubble. RPA70AB was placed close to the 3' junction where it interacts with XPG. We reoriented RPA70C to bind ssDNA near the 5' junction. The RPA70C/ssDNA was modeled by aligning the Ustilago maydis RPA/ssDNA structure (PDB ID: 4GOP) with the human trimer core structure (PDB ID: 1L1O). The orientation of RPA32D and RPA14 follows from the placement of the RPA70C module as they are all connected, forming the trimer core (70C/32D/14). To model XPA, we used the following structures: the cryo-EM TFIIH/XPA/DNA structure (PDB ID: 6RO4), the NMR structure of XPA/ERCC1 (PDB ID: 2JNW), and the human X-ray structure of RPA32C/Smarcal1 N-terminus (PDB ID: 4MQV). The XPA N-terminal extension (residues 1-103), which includes the RPA32C binding helix (residues 22-40), and the C-terminal extension (beta-domain) (residues 235-273) lacked known structural homologues and were modeled using AlphaFold2. The beta-domain was fitted into the TFIIH/XPA/DNA density. To position XPA's N-terminal helix (residues 22-40) we used the X-ray structure of RPA32C/Smarcal1 N-terminus. To assemble the complete PInC model, we also modelled loop regions of TFIIH's core subunits (XPB, XPD, p44, p34, and p52) into the TFIIH/XPA/DNA density.

5.5. Software	- AlphaFold2 (version Not available) - Modeller (version 10.4) - Clustal Omega (version Not available) - Coot (version 0.9.8.92) - Phenix (version 1.20.1) - UCSF Chimera (version 1.18)