Mechanism of mammalian transcriptional repression by noncoding RNA

Katarína Tlučková, Anita Salmazo, and Carrie Bernecky

Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria.

Abstract

Transcription by RNA polymerase II (Pol II) can be repressed by noncoding RNA, including the human RNA Alu. However, the mechanism by which endogenous RNAs repress transcription remains unclear. Here we present cryo-electron microscopy structures of Pol II bound to Alu RNA, which reveal that Alu RNA mimics how DNA and RNA bind to Pol II during transcription elongation. Further, we show how domains of the general transcription factor TFIIF affect complex dynamics and control repressive activity. Together, we reveal how a non-coding RNA can regulate mammalian gene expression.

Main text

RNA polymerase II (Pol II), the 12-subunit eukaryotic enzyme that generates mRNA, is a focal point for the regulation of transcription. The activity of Pol II is tightly regulated throughout the transcription cycle by association with various accessory factors, including single proteins, multi-subunit protein complexes, and RNAs[1]. Whereas protein regulators of transcription have been extensively studied, RNA-based mechanisms for regulation remain poorly understood.

Human Alu RNA has been identified as a natural inhibitor of transcription, blocking transcription of specific genes during heat shock[2]. Alu RNA is transcribed by Pol III as single units from short interspersed elements (SINEs), abundant genomic repeats. These free Alu RNAs are present at low levels under normal conditions, but levels increase in response to cellular stress[3]. Elevated levels of Alu RNA are able to repress Pol II transcription, and after heat shock, Alu RNA can be found at the promoters of repressed genes, consistent with the direct binding to Pol II observed *in vitro*[2]. Biochemical analysis revealed that Alu RNA binds directly to two molecules Pol II via independent interactions with the two halves of Alu RNA, Alu-left arm (Alu-LA, also known as scAlu) and Alu-right arm (Alu-RA)[2] (Fig 1a). Although both Alu-LA and Alu-RA bind to Pol II with high affinity, only Alu-RA is able to inhibit transcription in the presence of the general transcription factor TFIIF. It has been shown by low (~25 Å) resolution cryo-electron microscopy (cryo-EM) reconstructions[4] that both Alu-LA and Alu-RA localize to the Pol II DNA-binding cleft. Because the RNA secondary structure could not be resolved, it remains unknown how Alu RNAs are able to form high-affinity contacts with Pol

II, thus allowing them to repress within an endogenous context. Additionally, the mechanism by which the repression of only one of these very similar Alu RNA halves is affected by TFIIF remains to be elucidated.

In order to investigate the principles determining Pol II inhibitory activity of Alu RNA, we reconstituted complexes of mammalian Pol II bound to *in vitro* transcribed and refolded Alu-LA and Alu-RA (Extended Data Fig. 1). Our initial attempts at cryo-EM sample preparation using holey carbon or thin carbon-coated EM grids revealed that the samples displayed biased particle orientations, hindering high resolution analysis. Mild BS3 or glutaraldehyde crosslinking, which improved particle orientations for the Pol II elongation complex (EC)[5], resulted in dissociation of Alu RNA. Preparation of sample using graphene oxide-coated holey carbon grids[6] resulted in less biased particle orientations and allowed structure determination of Pol II-Alu-LA and Pol II-Alu-RA to nominal resolutions of 2.4 Å and 2.5 Å. For the Pol II-Alu-LA complex, the RNA could be resolved to approximately 5-8 Å. Similarly, Alu-RA within the Pol II-Alu-RA complex was resolved to approximately 5-8 Å. In both complexes, the Pol II clamp and stalk domains could be well resolved with further classification (Extended Data Figs. 2c and 3c).

The resolution of the RNA precluded building of an atomic model, but did reveal helical structure and the binding location of Alu RNA. Both Alu-LA and Alu-RA bind in the Pol II DNA binding cleft, adopting a conformation surprisingly similar to that of DNA and RNA within the Pol II EC (Fig 1b,c). Comparison to the Pol II EC revealed helical Alu RNA density overlapping the positions of the DNA-RNA hybrid, specific to transcribing complexes, and downstream DNA. Additional density partially overlapped the positions of upstream DNA and exiting RNA within the Pol II EC. Unlike in the EC, the Pol II clamp domain did not close over the nucleic acids in a stable conformation. Fork loop 2 within the DNA binding cleft, which interacts with the melted nontemplate DNA strand in the EC, was rearranged from the conformation observed in the EC to that observed in Pol II lacking DNA or RNA[5], consistent with the absence of a melted nucleic acid strand in the Alu RNA structures (Extended Data Fig. 3g). Altogether, despite Alu RNAs adopting an EC-like conformation, the Pol II enzyme was observed in an inactive state, consistent with the repressive activity of Alu RNA.

If Alu RNA resolution was restricted by intrinsic flexibility, we reasoned that a shorter RNA construct could overcome that limitation. To that end, we prepared a minimal version of Alu-RA (Fig. 1a) that was previously shown to be sufficient for both Pol II binding and repression[2]. We solved the structure of the Pol II-minimal Alu-RA complex to a nominal resolution of 3.1 Å, and observed that the quality of the RNA density was greatly improved (Fig. 2a). Surprisingly, 3D classification revealed the presence of an alternative RNA conformation, in which individual bases could be resolved, and an Alu-RA-like RNA conformation, still resolved at intermediate resolution (Fig. 2b, Extended Data Fig. 4). A model of Pol II and the well-resolved region of minimal Alu-RA in the alternative conformation was generated using a combination of *de novo* model building with DeepTracer[7] and secondary structure-based manual building[8] (Fig. 2c). The alternative conformation even more closely mimicked elongation complex nucleic acids (Fig. 2d,e). Comparison of the Pol II-minimal Alu-RA structure to that of *S. cerevisiae* Pol II bound to Fc aptamer[9], a synthetic RNA inhibitor, revealed that although both RNAs bind to a region overlapping that bound by the DNA-RNA hybrid, only minimal Alu-RA forms essentially the same interactions as those found within the elongation complex (Extended Data Fig. 5a,b).

Analogous to Alu RNA, bacterial 6S RNA is a natural noncoding RNA that directly binds the bacterial RNA polymerase and inhibits transcription from σ70-dependent housekeeping promoters. A previous near-atomic structure of *E. coli* RNA polymerase in complex with 6S RNA showed that this RNA adopted a conformation mimicking that of open promoter DNA[10]. This contrasts with our observation that Alu RNA mimics elongation complex nucleic acids (Fig. 1c). Previous results have shown the importance of a loosely structured region within Alu-RA for repressive activity[2] (Fig. 1a). We conclude that this region is important not as a mimic of open promoter DNA (Extended Data Fig. 5c), but rather allows the RNA to bend and adopt a conformation matching that of the elongation complex DNA-RNA hybrid and downstream DNA. This is consistent with previous observations that, unlike bacterial promoter complexes which can form a stable complex with just RNA polymerase and a sigma factor[11,12], eukaryotic initiation complexes require a large number of transcription factors to stabilize the interaction with DNA[13] and are dynamic[14]. Promoter complexes lacking accessory factors or a nascent RNA chain are unstable[15]. Thus, the more stable elongation complex[15-17] would be an effective mammalian intermediate to mimic.

Although Alu-LA and Alu-RA form similar structures, previous studies have shown that Alu-LA is more labile in the presence of TFIIF[18]. To investigate these differences in Pol II-Alu RNA complex stability, we first measured binding affinities and dissociation rates of Alu-LA and Alu-RA to Pol II using fluorescence anisotropy. We observed similarly high affinities for Alu-LA and Alu-RA when investigated in the presence of 150 mM monovalent salt, with the non-repressive Alu-LA binding with slightly higher affinity (Fig. 3a). The similar affinities were consistent with previous reports in which the complex was investigated in low salt conditions[2], despite a decrease of the apparent binding strength by approximately 10-fold. We then investigated the stability of the complexes in the presence of unlabeled competitor RNA, and found that Pol II-Alu-LA and Pol II-Alu-RA complexes both exhibited half-lives of approximately 15 minutes, significantly shorter than the approximately 90 minute half-lives observed in low salt conditions[2] (Fig. 3b). The binding and dissociation assays were repeated in the presence of TFIIF. TFIIF markedly reduced the half-life of Pol II-Alu-LA, whereas the half-life of Pol II-Alu-RA was only slightly affected. These results are consistent with previous work[2] and the observation that Alu-RA inhibits transcription more strongly than Alu-LA in the presence of TFIIF[2] (Fig. 3c). We propose that without the loose region of secondary structure shown to be important for Alu-RA transcription inhibition, Alu-LA is less stable in the bent conformation required to bind in an EC-like conformation, and therefore more labile and more susceptible to dissociation from Pol II bound to TFIIF.

Inspection of the Alu structures shows that TFIIF would not sterically clash with the resolved RNA elements (Fig. 3d)[19]. TFIIF has been shown to inhibit nonspecific DNA binding by Pol II, with the small subunit of TFIIF implicated as particularly important [20] It has also been suggested that a negatively charged region of the large subunit of TFIIF may be responsible for repelling non-specific DNA from the active site cleft[21]. To investigate how these regions of TFIIF may differentially affect Alu-LA and Alu-RA, we performed dissociation and transcription assays in the presence of TFIIF or TFIIF mutants (Fig. 3b,c). A TFIIF mutant within the charged region helix of the large subunit, TFIIF (W164A), known to be defective in transcription initiation but maintaining both domains thought to be important in the ability of TFIIF to inhibit nonspecific DNA binding[22], behaved essentially as wild-type TFIIF in both a dissociation and a transcription inhibition assay. We next investigated the effect of the winged helix domain of the small TFIIF subunit, a domain that binds upstream promoter DNA within the transcription initiation complex[23] and is important for transcription initiation activity[24]. TFIIF (RAP30ΔC) was unable to induce Pol II-Alu complex dissociation or to relieve transcriptional repression by either Alu construct. Lastly, we tested a TFIIF mutant lacking the C-terminus of the largest subunit, TFIIF (RAP74ΔC). The deleted region contained the negatively charged region and a winged helix domain. Unexpectedly, this mutant was able to enhance the Pol II dissociation of Alu-RA. We hypothesize that the negatively charged region may transiently bind to the positively-charged small subunit winged helix, reducing its nucleic acid binding capacity in the context of wild-type TFIIF. In the presence of TFIIF (RAP74ΔC), Alu-RA transcriptional repression was partially relieved, consistent with its decreased Pol II-bound half-life. In contrast, Alu-LA transcriptional repression activity increased slightly, suggesting that the C-terminal region may play an additional role in relieving the repressive activity of this Alu construct. Altogether, these results suggest that the winged helix of the TFIIF small subunit is important for regulating the stability of Pol II-Alu complexes, whereas the C-terminus of the large subunit, likely through electrostatic effects, contributes to the ability of TFIIF to relieve transcriptional repression by the Alu construct lacking the loosely structured region (Fig. 3e).

In summary, we found that Alu RNA is able to tightly bind Pol II via mimicry of elongation complex nucleic acids, in contrast to the promoter-like mimicry previously found in the bacterial system[10]. Although both halves of Alu RNA adopt the same elongation complex-like structure when bound to Pol II, transcriptional repressive activity of Alu RNA relies on the TFIIF-dependent dynamics of the complexes. We propose that elongation complex mimicry and dynamics-dependent activity may be a general mechanism of structured noncoding RNA in mammalian transcriptional regulation.

Methods

Cloning

The full length Alu sequence from cDNA clone TS 103[2,25] containing an A_{10} poly(A) tail was purchased as a gBlock (IDT) and cloned into a modified pSP64 plasmid (Promega) containing an autocleavable 3′-terminal hepatitis delta virus ribozyme[26]. Alu-LA (nucleotides 1-118) and Alu-RA (nucleotides 136-291) were PCR amplified and cloned into the modified pSP64 plasmid. A plasmid encoding wild type His_{10}-Arg_7-3C-SUMO-TFIIF was a gift from Dea Slade[27]. TFIIF mutants were created by round-the-horn site-directed mutagenesis. The TFIIF mutant W164A contains a single point mutation, W164A, in the charged region of GTF2F1 (RAP74)[28]. In TFIIF (RAP74ΔC), amino acids 181–517, encompassing the unstructured region and winged helix of GTF2F1 (RAP74), were deleted. In TFIIF (RAP30ΔC), amino acids 175-249, encompassing the GTF2F2 winged helix, were deleted.

RNA production and refolding

Plasmids were linearized with HindIII and used as templates for in vitro transcription using T7 RNAP (prepared “in-house”[29]). Alu RNAs were gel purified and ethanol precipitated. Ribozyme cleavage resulted in 3’ cyclic phosphates, which were removed by T4 PNK (NEB) treatment (100 mM MES-NaOH pH 5.5 at RT, 10 mM $MgCl_2$, 10 mM 2-Mercaptoethanol, 300 mM NaCl, 1U/µL RiboLock RNase Inhibitor (Thermo Scientific)) as previously described[30]. RNA was desalted using a PD10 column and ethanol precipitated. Purified RNA was refolded by heating to 95°C for 5 min in buffer as specified, followed by snap cooling on ice prior to each experiment.

RNA labeling

RNAs were 3’end-labeled with ATTO 488 hydrazide (ATTO-TEC). RNA was oxidized by adding fresh sodium periodate to a final concentration of 40 µM and incubated at 25°C for 90 min in the dark. Reactions were quenched by adding fresh sodium thiosulfate to a final concentration of 80 µM and incubated at 25°C for 25 min in the dark. A 30-fold molar excess of ATTO 488 hydrazide was added and the reaction was incubated for 4h at 25°C, protected from light. RNA was recovered by phenol-chloroform extraction and ethanol precipitation. The unreacted dye was removed using a Quick-RNA Kit (Zymo Research).

Protein preparation

S. scrofa Pol II was purified as previously described[5]. The porcine and human enzymes are 99.9% identical, with only 4 changes within the 4587-amino-acid sequence. TFIIF was recombinantly expressed in BL21 DE3-RIL *E. coli*. Cultures were grown to an OD_{600} of 0.9 and induced by addition of 0.5 mM IPTG at 37°C for 4h. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol, 1 mM DTT) plus protease inhibitors (1 mM PMSF, 2 mM benzamidine, 1 µM leupeptin and 2 µM pepstatin), and lysed by sonication. TFIIF was purified by affinity chromatography using a HisTrap HP column (Cytiva) equilibrated with lysis buffer. Protein was eluted with lysis buffer containing 200 mM imidazole, dialyzed (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT), and cleaved by 3C protease at 4°C overnight. Cleaved protein was purified using a HiTrap SP column equilibrated with ion exchange buffer (50 mM HEPES pH 7.5, 10% (v/v) glycerol, 2 mM DTT) containing 150 mM NaCl. Protein was eluted using a linear gradient of 0.15-1 M NaCl in ion exchange buffer and further purified using a MonoQ column equilibrated with ion exchange buffer containing 100 mM NaCl. Protein was eluted as before and subjected to size exclusion chromatography using a Superdex 200 10/300 column (TFIIF wild-type, W164A, and RAP30ΔC) or a Superdex 75 10/300 column (TFIIF RAP74ΔC) (Cytiva) equilibrated with 50 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol and 2 mM DTT. Purified TFIIF was concentrated using a Amicon Ultra-4 concentrator with a 30kDa cut-off (Amicon) and stored at -80 °C until use.

Mass photometry

Refolded RNA was diluted to 1 µM in 50 mM KCl binding buffer (20 mM HEPES pH 7.5, 50 mM KCl, 4 mM $MgCl_2$, 10 mM DTT). Pol II was diluted to 1 µM in 50 mM KCl binding buffer. The complex was prepared by mixing 2.5 µl diluted Pol II with 2.5 µl diluted RNA, followed by incubation at 30°C for 30 min. Samples were placed on ice and diluted 5X in 50 mM KCl binding buffer. Measurements were carried out using a Refeyn Two^{MP} mass photometer (Refeyn Ltd.). The instrument was focused with 10 µl buffer and 1 µl of sample was used for each measurement. Movies were collected for 1 min and processed with DiscoverMP software. Calibration of the instrument was performed prior each data collection session.

Cryo-EM sample preparation

Pol II was diluted in 150 mM NaCl polymerase buffer (5 mM HEPES pH 7.25, 150 mM NaCl, 10 µM $ZnCl_2$, 10 mM DTT) and mixed with a 2-fold excess of refolded RNA in 150 mM NaCl polymerase buffer. The complex was incubated at 30°C for

30 min, diluted to a salt concentration of 50 mM NaCl, and placed on ice. The final concentration of Pol II was 400 nM.

Four microliters of sample were applied to graphene oxide-coated Quantifoil R1.2/1.3 holey carbon grids[6] that had been glow discharged for 5 s (23 mA current, 7.0 x 10^{-1} mbar vacuum). Using a Vitrobot Mark IV set to 100% humidity and 4 °C, grids were blotted for 13 seconds and immediately plunge frozen in liquid ethane.

Cryo-electron microscopy

Grids were screened for particle density, particle orientation, and graphene oxide coverage using a Thermo Fisher Glacios transmission electron microscope (200 kV) equipped with a Falcon III direct electron detector. Grids were transferred to a Thermo Fisher Titan Krios G3i transmission electron microscope (300 kV) equipped with a Gatan K3 BioQuantum direct electron detector (energy slit width 10 eV) for data collection. Data sets were recorded using SerialEM 3.8 and EPU 2.11 (Pol II-Alu-RA), SerialEM 3.9 (Pol II-Alu-LA), and EPU 2.11 (Pol II-minimal Alu-RA). Micrographs were collected at a nominal magnification of ×105000, corresponding to a counting mode object scale pixel size of 0.835 Å. For Pol II-Alu-LA, the data set consisted of 16,740 micrographs collected using a defocus range of -0.2 to -2.0 µm with an electron exposure rate of 16.48 $e^-/Å^2/s$ and an exposure time 2.14 s. The total electron exposure was 40 $e^-/Å^2$ distributed over 40 frames. For Pol II-Alu-RA, the data set consisted of 29,262 micrographs collected using a defocus range of -0.2 to -2.2 µm with an electron exposure rate of 14.2 $e^-/Å^2/s$ and an average exposure time of 2.2 s. The total electron exposure was 40 $e^-/Å^2$ distributed over 40 frames. For Pol II-minimal Alu-RA, the data set consisted of 7,588 micrographs collected using a defocus range of -0.1 to -2.2 µm with an electron exposure rate of 22.7 $e^-/Å^2/s$ and an exposure time 1.81 s. The total electron exposure was 41 $e^-/Å^2$ distributed over 40 frames.

Micrographs were first processed with Warp[31] and visually screened to remove micrographs with poorly visible Thon rings or multiple layers of graphene oxide. Micrographs were motion corrected and dose weighted using RELION 3.1[32], followed by CTF estimation using CTFFIND4[33]. Particle coordinates were selected with Topaz[34] using a very permissive threshold to ensure all particles were identified. Data sets were subsequently cleaned using a combination of 2D and 3D classification in RELION 3.1 and 2D classification in CryoSPARC[35] (Extended Data Figs. 2, 3, and 4). Final reconstructions were obtained using 3D refinement with defocus refinement, beam tilt refinement, and particle polishing in RELION 3.1 and a final binned pixel size of 1.0 Å. Local resolution was calculated in CryoSPARC using an FSC cutoff of 0.5. Densities displaying well-resolved clamp and stalk domains could be obtained after additional 3D classification with appropriate masks (Extended Data Figs. 2 and 3).

Modeling and refinement

A model for the Pol II core lacking the stalk and clamp domains, generated from the Pol II DSIF-EC structure (PDB ID 5OIK)[36], was rigid body fit into the Pol II-Alu-RA map using UCSF Chimera[37]. The conformation of fork loop 2 (RBP2 487-499) was manually adjusted in Coot[38], and the resulting model was adjusted using Isolde[39]. The model was real space refined in Phenix[40] using Isolde-suggested parameters (global minimization with reference restraints and ADP refinement). The minimal Alu-RA RNA was *de novo* modeled into an initial Pol II-minimal Alu-RA map (prior to classification of the two RNA conformations) using DeepTracer[7]. Using the final classified, b-factor-sharpened minimal Alu-RA alternative conformation map, the model was adjusted in Coot[38] and the sequence register was assigned, guided by the minimum free energy RNAfold[8] secondary structure prediction generated using energy parameters rescaled to 4 °C, as energy parameters at 37°C yielded the canonical Alu-RA conformation. Base G52 was not built, despite potential density for a base stacked on residue RBP1 Y859, due to density quality. Clamp and adjusted core Pol II models were rigid body fitted into the Pol II-minimal Alu-RA alternative conformation map, regions lacking any density in the unsharpened EM map were deleted, and the combined Pol II-RNA model was further adjusted in Isolde. The model was real space refined in Phenix[40] using Isolde-suggested parameters (global minimization with reference restraints and ADP refinement). Densities and models were visualized using UCSF ChimeraX[41]. Difference densities for Pol II-Alu-LA and Pol II-Alu-RA reconstructions were visualized after subtraction of individually fitted Pol II core, clamp, and stalk domains. For visualization purposes, the Pol II-minimal Alu-RA alternative conformation map was post-processed with DeepEMHancer[42].

Fluorescence anisotropy binding and dissociation experiments

Pol II and ATTO-labeled RNA were diluted in 150 mM NaCl binding buffer. A complex of Alu RNA and Pol II was prepared using 5 nM RNA and increasing concentrations of Pol II (0-1.28 μM). Reactions were prepared in a final volume of 20 μl (20 mM HEPES pH 7.5, 150 mM NaCl, 10 μM $ZnCl_2$, 4 mM $MgCl_2$, 10 mM DTT) and incubated at 30 °C for 30 min. All experiments were performed in triplicate. Fluorescence anisotropy was measured in 384-well plates and data were collected using a Plate Reader Synergy H1-MF (Bio-TEK). Data were analyzed in GraphPad Prism 9 using a single site quadratic binding equation accounting for ligand depletion[43]. Values represent the mean and error bars represent the standard deviation. Significance level (p-value) was calculated using an unpaired, two-tailed t-test.

For dissociation experiments, ATTO-labeled RNA and Pol II were diluted in 150 mM NaCl binding buffer. The complex was prepared at a final concentration of 100 nM Pol II and 5 nM RNA, then incubated at 30 °C for 30 min. Immediately prior to the dissociation measurement, a 100-fold excess of unlabeled RNA over labeled RNA was added and fluorescence anisotropy was measured at 30 °C for 3h at 3 min intervals. Each sample was prepared in six replicates. Data were analyzed in GraphPad Prism 9 using the built-in dissociation kinetics model (one phase exponential decay).

Transcription and inhibition assays

Pol II was diluted in transcription buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 5 μM $ZnCl_2$, 3 mM $MgCl_2$, 4 (v/v) % glycerol, 1 mM DTT) and mixed with 2.4 pmol of a tailed DNA template (template strand sequence 5’-ACAAATTACTGGGAAGTCGACTATGCAATACAGGCATCATTTGATCAAGCTCAAGTACTTAATCATAACCATA-3’, nontemplate strand sequence 5’-TAAGTACTTGAGCTTGATCAAATGATGCCTGTATTGCATAGTCGACTTCCCAGTAATTTGT-3’) and bovine serum albumin (20 μg/ml). Reactions were incubated for 10 min at 30°C. The final concentration of the Pol II was 0.24 μM. An NTP mix containing radioactive CTP (625 μM ATP, GTP, UTP; 25 μM CTP and 25 μM [α-32P] CTP) was added to the reaction and incubated for 10 min at 37°C. Transcription was stopped by adding an equal volume of stop buffer (8M Urea, 20 mM EDTA), incubated for 5 min at 95°C, and snap cooled on ice. The reaction was incubated with Proteinase K (0.2 mg/ml) for 20 min at 37°C, then heat inactivated for 5 min at 95°C. RNA was resolved on a denaturing gel (20% acrylamide, 8M Urea).

For inhibition assays, folded Alu RNA was preincubated with Pol II for 30 min at 30°C (protein:RNA molar ratio of 1:2). DNA template was added and transcription was carried out as described above. For reactions containing TFIIF, Pol II was preincubated with TFIIF or TFIIF mutants (Pol II:TFIIF molar ratio of 1:4) for 1h at 30°C, then Alu RNA was added RNA to the reaction. Assays were carried out as described above.
For detection of radioactivity, gels were exposed to a storage phosphor screen overnight at 4°C and imaged with a Typhoon 9400 scanner (Cytiva). Data were analyzed using ImageQuant TL 10.2 analysis software (Cytiva).

References

1. Schier, A.C. & Taatjes, D.J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev* **34**, 465-488 (2020).
2. Mariner, P.D. et al. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* **29**, 499-509 (2008).
3. Liu, W.M., Chu, W.M., Choudary, P.V. & Schmid, C.W. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* **23**, 1758-65 (1995).
4. Kassube, S.A. et al. Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II. *J Mol Biol* **425**, 3639-48 (2013).

5. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551-4 (2016).
6. Palovcak, E. et al. A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J Struct Biol* **204**, 80-84 (2018).
7. Pfab, J., Phan, N.M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc Natl Acad Sci U S A* **118**(2021).
8. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
9. Kettenberger, H. et al. Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. *Nat Struct Mol Biol* **13**, 44-8 (2006).
10. Chen, J. et al. 6S RNA Mimics B-Form DNA to Regulate Escherichia coli RNA Polymerase. *Mol Cell* **68**, 388-397 e6 (2017).
11. Roe, J.H., Burgess, R.R. & Record, M.T., Jr. Kinetics and mechanism of the interaction of Escherichia coli RNA polymerase with the lambda PR promoter. *J Mol Biol* **176**, 495-522 (1984).
12. Saecker, R.M. et al. Structural origins of Escherichia coli RNA polymerase open promoter complex stability. *Proc Natl Acad Sci U S A* **118**(2021).
13. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P.A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-61 (1989).
14. Treutlein, B. et al. Dynamic architecture of a minimal RNA polymerase II open promoter complex. *Mol Cell* **46**, 136-46 (2012).
15. Cai, H. & Luse, D.S. Transcription initiation by RNA polymerase II in vitro. Properties of preinitiation, initiation, and elongation complexes. *J Biol Chem* **262**, 298-304 (1987).
16. Cheung, A.C., Sainsbury, S. & Cramer, P. Structural basis of initial RNA polymerase II transcription. *EMBO J* **30**, 4755-63 (2011).
17. Liu, X., Bushnell, D.A., Silva, D.A., Huang, X. & Kornberg, R.D. Initiation complex structure and promoter proofreading. *Science* **333**, 633-7 (2011).
18. Wagner, S.D., Kugel, J.F. & Goodrich, J.A. TFIIF facilitates dissociation of RNA polymerase II from noncoding RNAs that lack a repression domain. *Mol Cell Biol* **30**, 91-7 (2010).
19. Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**(2021).
20. Killeen, M.T. & Greenblatt, J.F. The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. *Mol Cell Biol* **12**, 30-7 (1992).
21. Chen, Z.A. et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* **29**, 717-26 (2010).
22. 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**, 7377-87 (1999).
23. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-65 (2016).

24. Tan, S., Conaway, R.C. & Conaway, J.W. Dissection of transcription factor TFIIF functional domains required for initiation and elongation. *Proc Natl Acad Sci U S A* **92**, 6042-6 (1995).
25. Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. & Deininger, P.L. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J Mol Biol* **271**, 222-34 (1997).
26. Martinez-Rucobo, F.W. et al. Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol Cell* **58**, 1079-89 (2015).
27. Appel, L.M. et al. PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat Commun* **12**, 6078 (2021).
28. 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**, 3583-95 (2005).
29. Ichetovkin, I.E., Abramochkin, G. & Shrader, T.E. Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J Biol Chem* **272**, 33009-14 (1997).
30. Schurer, H., Lang, K., Schuster, J. & Morl, M. A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res* **30**, e56 (2002).
31. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019).
32. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
33. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).
34. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160 (2019).
35. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
36. Bernecky, C., Plitzko, J.M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* **24**, 809-815 (2017).
37. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
38. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
39. Croll, T.I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018).
40. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
41. Pettersen, E.F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).

42. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874 (2021).
43. Vos, S.M. et al. Architecture and RNA binding of the human negative elongation factor. *Elife* **5**(2016).
44. Lei, L., Ren, D. & Burton, Z.F. The RAP74 subunit of human transcription factor IIF has similar roles in initiation and elongation. *Mol Cell Biol* **19**, 8372-82 (1999).
45. Fang, S.M. & Burton, Z.F. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J Biol Chem* **271**, 11703-9 (1996).

Acknowledgements

We thank B. Kaczmarek and other members of the Bernecky lab for helpful discussions. We thank V.-V. Hodirnau for SerialEM data collection and support with EPU data collection. We thank D. Slade for the wild type TFIIF expression plasmid. We thank N. Thompson and R. Burgess for the 8WG16 hybridoma cell line. We thank C. Plaschka and M. Loose for critical reading of the manuscript. This work was supported by Austrian Science Fund (FWF) grant P34185. This research was further supported by the Scientific Service Units (SSU) of IST Austria through resources provided by the Lab Support Facility (LSF), Electron Microscopy Facility (EMF), Scientific Computing (SciComp), and the Preclinical Facility (PCF).

Author contributions

K.T. performed the experiments and analyzed the data. A.T.S. assisted with protein purification. C.B. designed and supervised research. K.T. and C.B. processed cryo-EM data and prepared the manuscript.

Data availability

Cryo-EM maps for Pol II-Alu-LA, Pol II-Alu-RA, Pol II-minimal-Alu-RA canonical conformation, and Pol II-minimal-Alu-RA alternative conformation were deposited to the EM Data Bank under the accession codes EMD-18367, EMD-18375, EMD- 18371, and EMD-18376. Model coordinates for the Pol II-AluRA polymerase core and Pol II-minimal Alu-RA complex were deposited to the PDBe under the accession codes 8QEP and 8QEQ.

Competing interests

The authors declare no competing interests.

Corresponding author

Correspondence to carrie.bernecky@ist.ac.at.

Figures

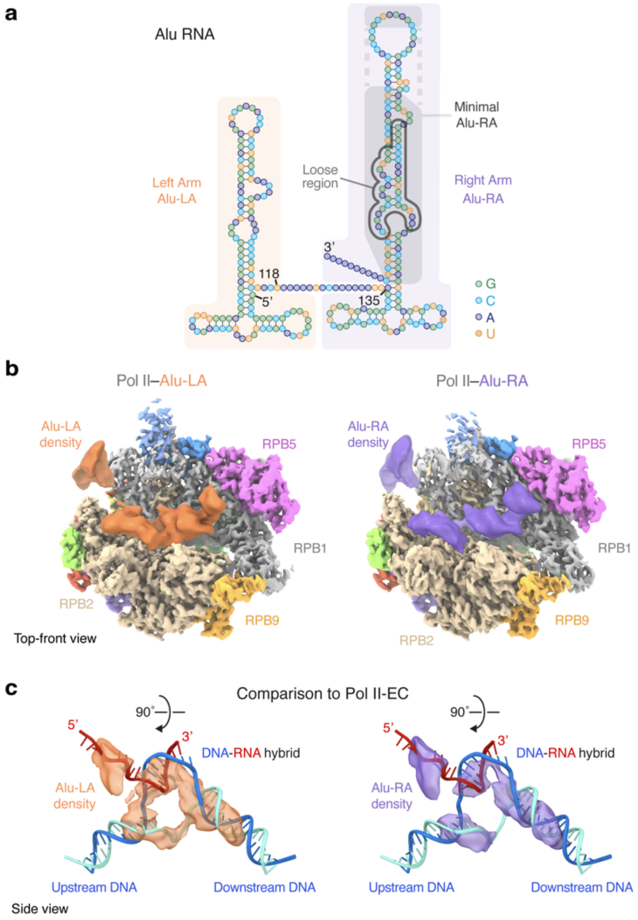


Figure 1 Structural analysis of Pol II-Alu-LA and -Alu-RA complexes. **a**, Schematic showing the secondary structure and domain organization of Alu RNA. Alu-LA (orange shading), Alu-RA (purple shading), Alu-RA minimal region (gray shading), and Alu-RA loose region (gray outline) are indicated. Guanosine, green; cytidine, cyan; adenosine, blue; uridine, orange. **b**, Top-front view of the Pol II-Alu-LA and -Alu-RA reconstructions (not b-factor-sharpened), with 8-Å-filtered Alu-LA (orange) and Alu-RA (purple) difference densities overlaid. Pol II is colored by subunit. **c**, Side view of 8-Å-filtered Alu-LA and Alu-RA densities compared to the path of nucleic acids in a transcription elongation complex (PDB ID 5OIK)[36].

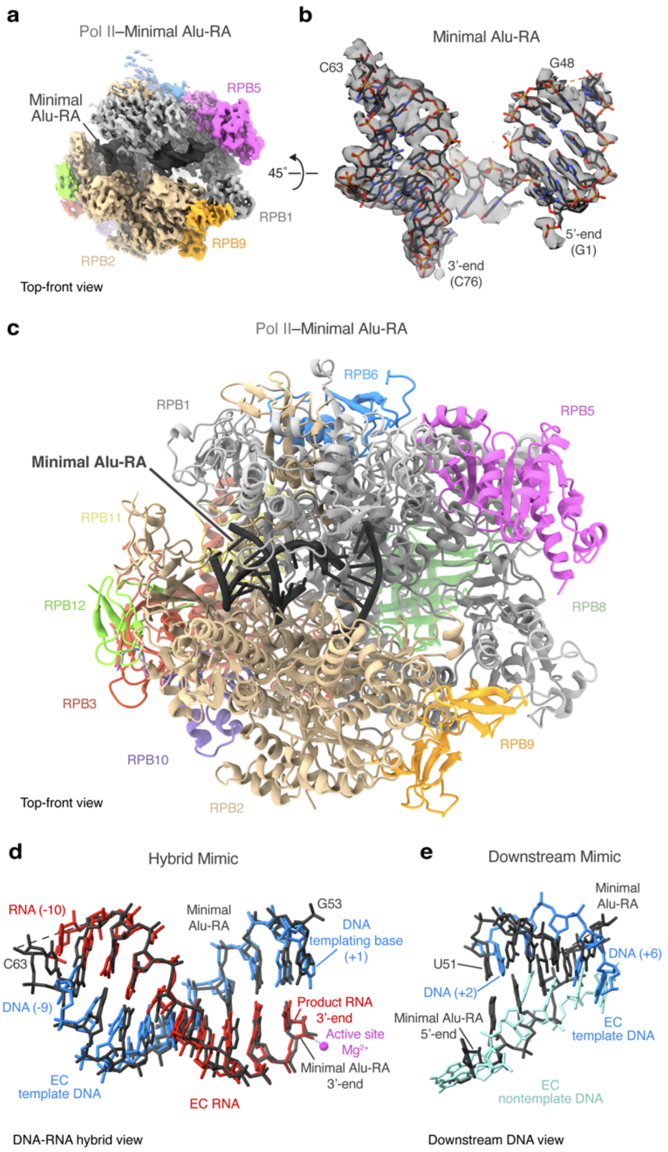


Figure 2 Structure of Pol II-minimal Alu-RA. Coloring as in Fig. 1. **a**, Top-front view of the locally-filtered Pol II-minimal Alu-RA alternative conformation density. **b**, A rotated view of the modeled region of minimal Alu-RA overlaid with the EM density (DeepEMhancer post-processed for visualization). **c**, Top-front view of the Pol II-minimal Alu-RA structure. **d**, A view of the DNA-RNA hybrid-like region of minimal Alu-RA and the same region of a transcription elongation complex (PDB ID 5OIK)[36]. **e**, As in **d**, but focused on the downstream DNA region.

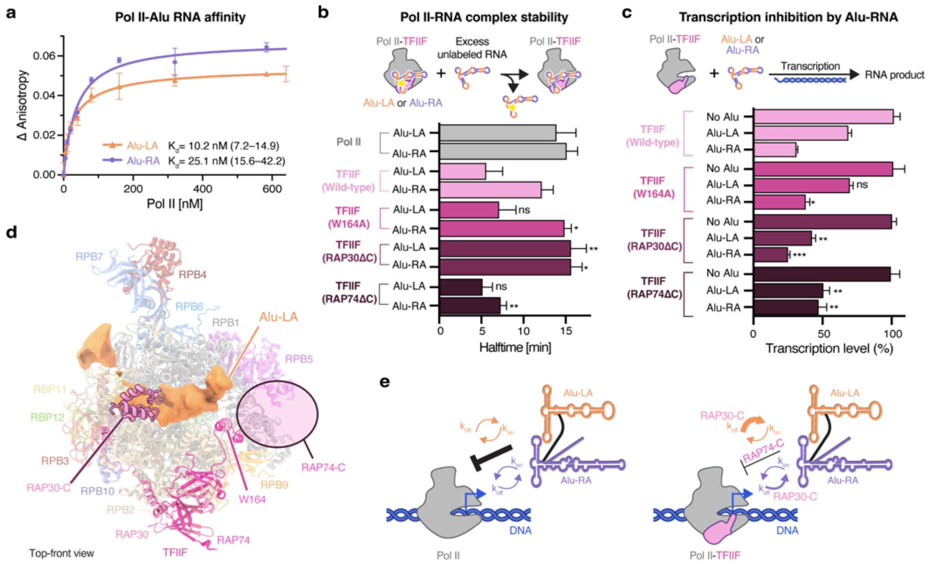


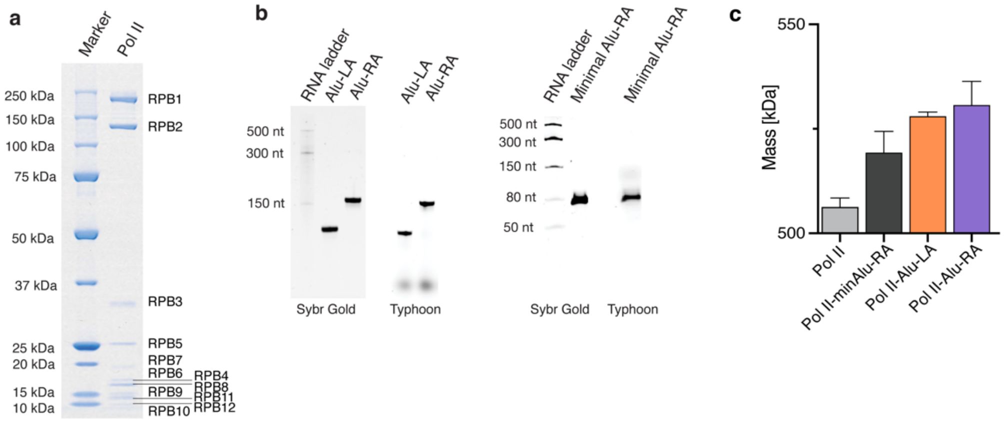
Figure 3 Effects of TFIIF and its domains on dynamics and activity of Pol II-Alu RNA complexes. **a**, Fluorescence anisotropy experiments reveal Alu-LA and Alu-RA binding affinity to Pol II. **b**, Fluorescence anisotropy dissociation experiments reveal the half-life of Pol II-Alu-LA and Pol II-Alu-RA complexes either in the absence of TFIIF, or the presence of wild-type TFIIF or the indicated TFIIF mutants. Achieved significance level relative to the wild-type TFIIF condition: ***, $p \leq 0.001$; **, $p \leq 0.01$; *, $p \leq 0.05$; ns, $p > 0.05$. **c**, Transcription assays carried out in the presence of the indicated TFIIF variant and either no RNA, Alu-LA, or Alu-RA. Asterisks represent achieved significance level relative to the wild-type TFIIF condition as in **b**. **d**, Alu-LA density overlaid on Pol II-TFIIF as found within a transcription initiation open complex (PDB ID 5IYB)[23]. Pol II subunits are shown as semi-transparent ribbons. The mutated regions of the TFIIF constructs used in **b** and **c** are indicated. **e**, Schematic representation of the dynamics and activity of Alu RNA halves in the absence and presence of TFIIF.

Table 1 Cryo-EM data collection, refinement, and validation statistics.

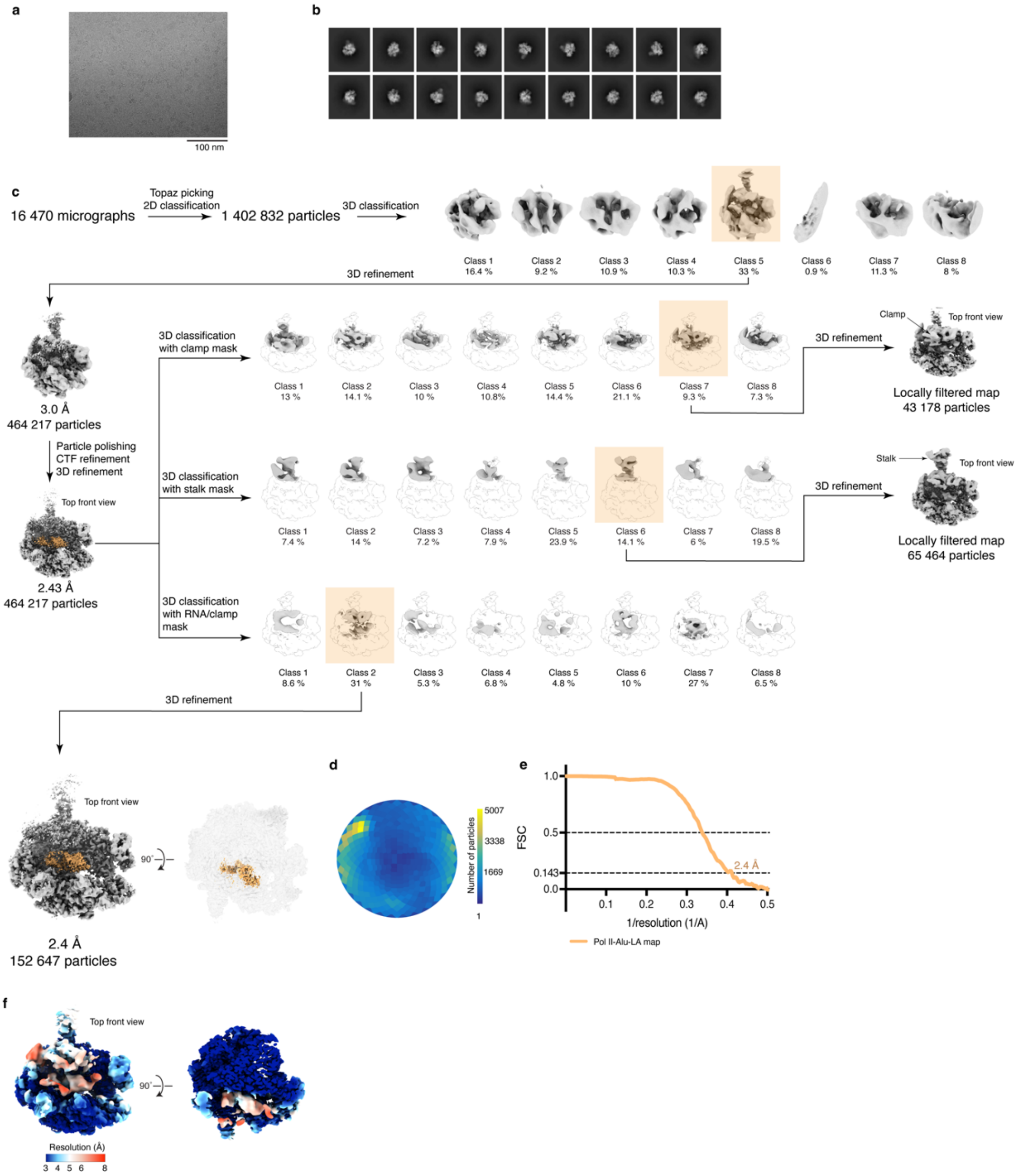
	Pol II-Alu-LA (EMDB-18367)	Pol II-Alu-RA (EMDB-18375) (PDB 8QEP)	Pol II-minimal Alu-RA, canonical conformation (EMDB-18371)	Pol II-minimal Alu-RA, alternative conformation (EMDB-18376) (PDB 8QEQ)
Data collection and processing				
Magnification	105,000x	105,000x	105,000x	105,000x
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^{2})	40	40	41	41
Defocus range (µm)	-0.2 to -2.0	-0.2 to -2.2	-0.1 to -2.2	-0.1 to -2.2
Pixel size (Å)	0.835	0.835	0.835	0.835
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	1,402,832	1,984,654	722.886	722.886
Final particle images (no.)	152.567	857.265	72.783	90.778
Map resolution (Å)	2.4	2.5	3.2	3.1
FSC threshold				
Map resolution range (Å)	2.3-8.0	2.3-8.0	2.7-8.0	2.7-8.0
Refinement				
Initial model used (PDB code)	-	5OIK	-	5OIK
Model resolution (Å)	-	2.6	-	3.2
FSC threshold		0.5		0.5
Map sharpening *B* factor (Å^{2})	-40.63	-50.93	-46.12	-39.56
Model composition	-		-	
Non-hydrogen atoms		25496		29257
Protein residues		3177		3575
Ligands		4		8
B factors (Å^{2})	-		-	
Protein		39.65		49.14
Ligand		70.50		92.84
R.m.s. deviations	-		-	
Bond lengths (Å)		0.003		0.004
Bond angles (°)		0.725		0.748

Validation	-		-	
MolProbity score		1.40		1.44
Clashscore		2.24		2.40
Poor rotamers (%)		1.63		1.39
Ramachandran plot	-		-	
Favored (%)		96.40		95.50
Allowed (%)		3.54		4.22
Disallowed (%)		0.06		0.28

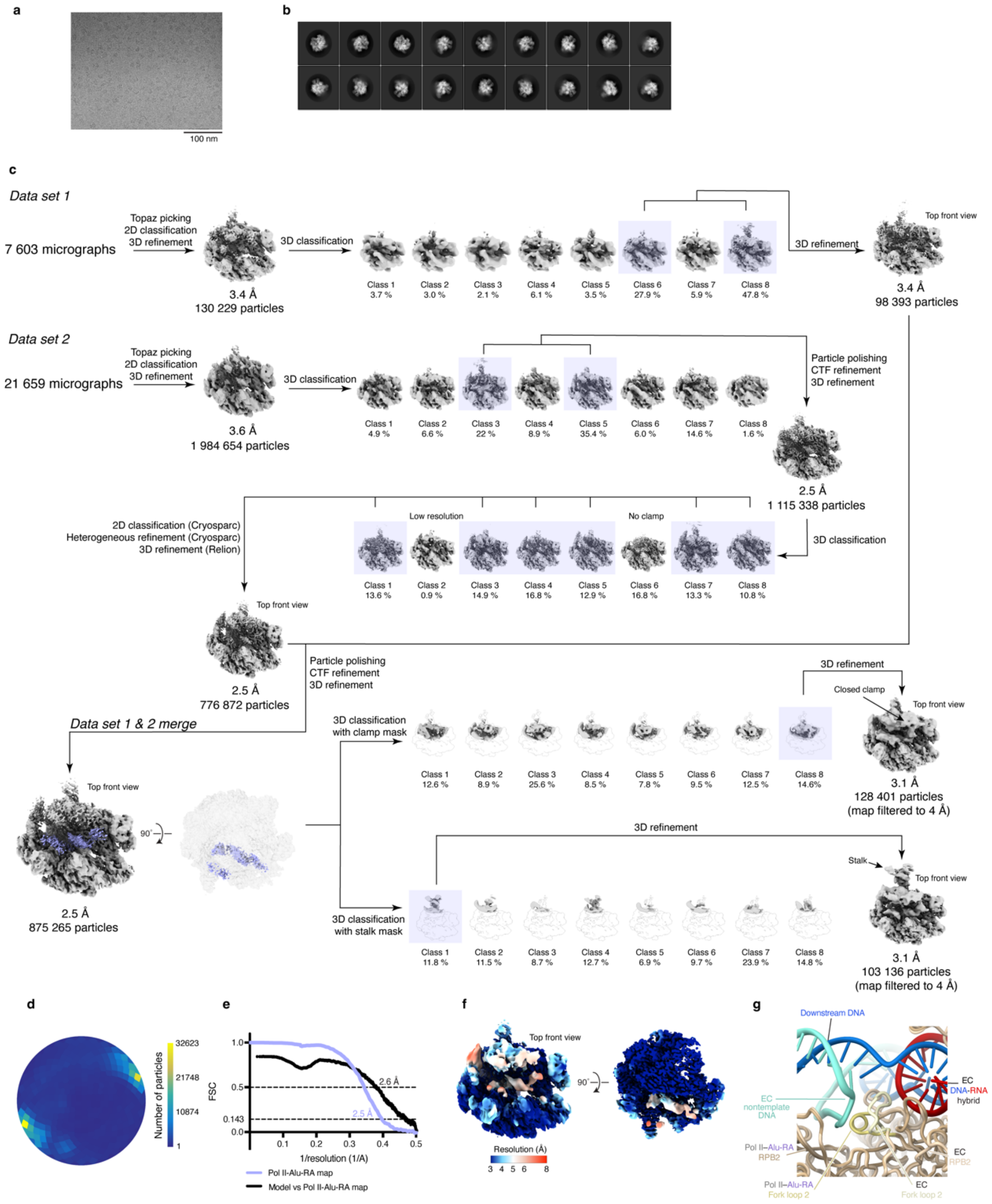
Extended data



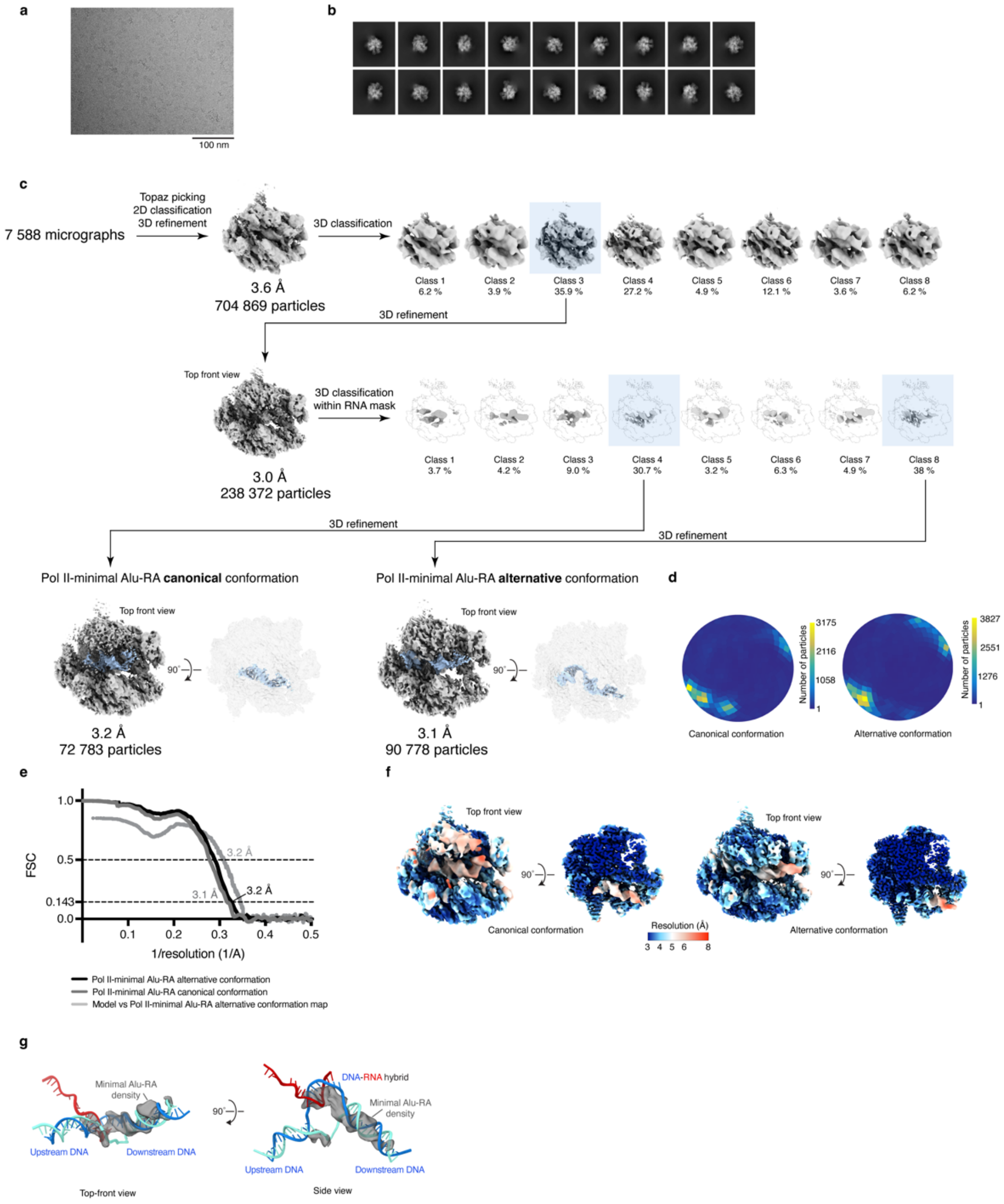
Extended Data Figure 1 Preparation of Pol II-Alu RNA complexes. **a**, SDS-PAGE analysis (Coomassie) of purified Pol II **b**, Denaturing urea-PAGE analysis of purified Alu RNAs. Unlabeled RNAs (left) were visualized by SYBR Gold staining, and fluorescently labeled RNAs were visualized using a Typhoon RGB scanner. **c**, Mass photometry analysis of Pol II-Alu complexes. The mass of the monomeric Pol II mass photometry peak was measured in triplicate. Error bars represent the standard deviation.



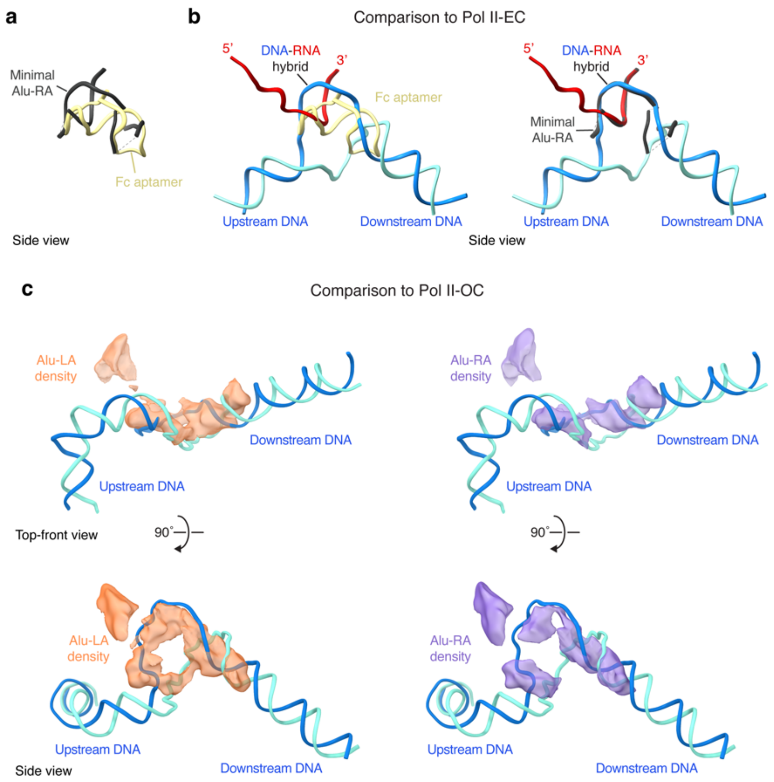
Extended Data Figure 2 Cryo-EM analysis of Pol II-Alu-LA. **a**, Cryo-EM micrograph representative field of view, -1.9 μm defocus. **b**, Representative 2D classes. **c**, Processing tree outlining the steps taken to generate the Pol II-Alu-LA reconstruction. Classes with well-resolved clamp and stalk domains are indicated. **d**, Angular distribution plot for the Pol II-Alu-LA reconstruction. **e**, Fourier shell correlation plot for the Pol II-Alu-LA reconstruction. **f**, Locally-filtered, non-b-factor-sharpened Pol II-Alu-LA reconstruction colored by estimated local resolution.



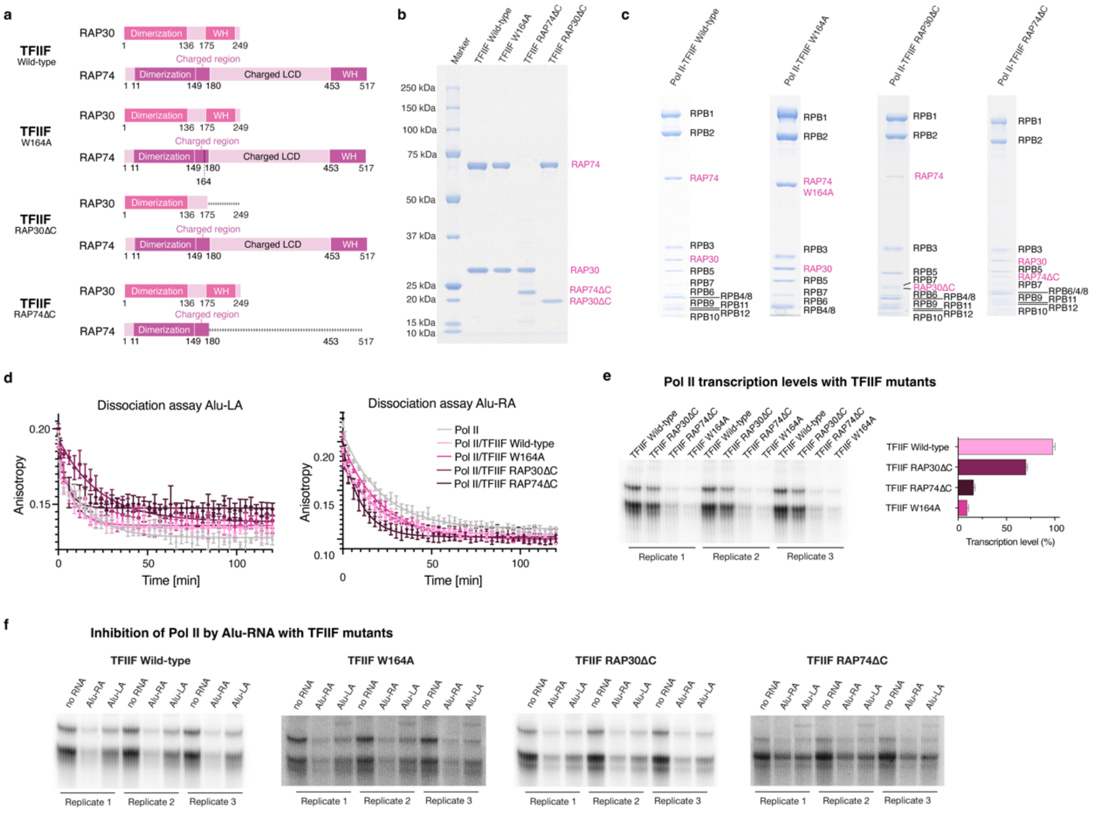
Extended Data Figure 3 Cryo-EM analysis of Pol II-Alu-RA. **a**, Cryo-EM micrograph representative field of view, -1.6 μm defocus. **b**, Representative 2D classes. **c**, Processing tree outlining the steps taken to generate the Pol II-Alu-RA reconstruction. Classes with well-resolved clamp and stalk domains are indicated. **d**, Angular distribution plot for the Pol II-Alu-RA reconstruction. **e**, Fourier shell correlation plot for the Pol II-Alu-RA reconstruction and the model-versus-map correlation for the Pol II-Alu-RA reconstruction and the Pol II core model. **f**, Locally-filtered, non-b-factor-sharpened Pol II-Alu-RA reconstruction colored by estimated local resolution. **g**, Comparison of the fork loop 2 conformation in the Pol II-Alu-RA structure and a transcription elongation complex structure (PDB ID 5OIK)[36]. Coloring is as follows: Pol II-Alu-RA: RBP2 tan, fork loop 2 yellow; EC: RPB2 beige, fork loop 2 light yellow, template DNA blue, nontemplate DNA cyan, RNA red.



Extended Data Figure 4 Cryo-EM analysis of Pol II-minimal Alu-RA. **a**, Cryo-EM micrograph representative field of view, -1.7 μm defocus. **b**, Representative 2D classes. **c**, Processing tree outlining the steps taken to generate the Pol II-minimal Alu-RA reconstructions. **d**, Angular distribution plot for the Pol II-minimal Alu-RA reconstructions. **e**, Fourier shell correlation plot for the Pol II-minimal Alu-RA reconstructions and the model-versus-map correlation for the Pol II-minimal Alu-RA alternative conformation reconstruction and the model. **f**, Locally-filtered Pol II-minimal Alu-RA reconstructions colored by estimated local resolution. The Pol II-minimal Alu-RA canonical conformation reconstruction was not b-factor sharpened. **g**, Top-front and side views of 8-Å-filtered Pol II-minimal Alu-RA canonical conformation density compared to the path of nucleic acids in a transcription elongation complex (PDB ID 5OIK)[36]. Minimal Alu-RA density is shown in semi-transparent gray. Compare to Fig. 1b,c.



Extended Data Figure 5 Comparison of Alu RNA and nucleic acids within other Pol II complexes. **a**, Comparison of the paths of the nucleic acids in the Pol II-minimal Alu-RA structure and the yeast Pol II-Fc aptamer structure (PDB ID 2B63)[9] after alignment on RPB2. **b**, Comparison of the paths of Fc aptamer and minimal Alu-RA and nucleic acids in a transcription elongation complex (PDB ID 5OIK)[36]. **c**, Top-front and side views of 8-Å-filtered Alu-LA and Alu-RA densities compared to the path of nucleic acids in a transcription initiation open promoter complex (PDB ID 5IYB)[23].



Extended Data Figure 6 Preparation and activity of transcription complexes with TFIIF and TFIIF mutants. **a**, Schematic detailing the TFIIF variants used in these studies. **b**, SDS-PAGE analysis of equimolar amounts of each TFIIF variant. **c**, SDS-PAGE analysis of size-exclusion-purified Pol II-TFIIF complexes for use in dissociation assays. **d**, Fluorescence anisotropy dissociation curves quantified in Fig. 3b. **e**, Transcription assays show that the TFIIF variants employed in this study have differential ability to stimulate Pol II activity, as previously described[44,45]. Left, Urea PAGE analysis and phosphorimaging; right, quantification. **f**, Urea-PAGE analysis of transcription assay quantified in Fig. 3c.