

RNA polymerase I structure and transcription regulation

Christoph Engel¹, Sarah Sainsbury¹, Alan C. Cheung¹, Dirk Kostrewa¹ & Patrick Cramer¹

Transcription of ribosomal RNA by RNA polymerase (Pol) I initiates ribosome biogenesis and regulates eukaryotic cell growth. The crystal structure of Pol I from the yeast *Saccharomyces cerevisiae* at 2.8 Å resolution reveals all 14 subunits of the 590-kilodalton enzyme, and shows differences to Pol II. An ‘expander’ element occupies the DNA template site and stabilizes an expanded active centre cleft with an unwound bridge helix. A ‘connector’ element invades the cleft of an adjacent polymerase and stabilizes an inactive polymerase dimer. The connector and expander must detach during Pol I activation to enable transcription initiation and cleft contraction by convergent movement of the polymerase ‘core’ and ‘shelf’ modules. Conversion between an inactive expanded and an active contracted polymerase state may generally underlie transcription. Regulatory factors can modulate the core–shelf interface that includes a ‘composite’ active site for RNA chain initiation, elongation, proofreading and termination.

Eukaryotic genes are transcribed by Pol I, II and III (ref. 1), which synthesize ribosomal, messenger and transfer RNA, respectively. Pol I synthesizes about 60% of cellular RNA by transcribing several copies of the rRNA gene. Each gene copy is transcribed by dozens of Pol I enzymes, generating microscopic ‘Christmas tree’ structures². Pol I activity is a key determinant for the level of ribosome components and controls cell growth³. Pol I deregulation occurs in many diseases including cancer⁴.

Two decades ago the shape of Pol I was revealed by electron microscopy^{5,6}, but the crystal structure of Pol I is lacking. The structure of Pol II was, however, solved for its 10-subunit core at the turn of the millennium⁷. Subsequent structures of Pol II complexes have elucidated the transcription mechanism⁸. Pol II contains a conserved active centre cleft with a catalytic aspartate loop, a bridge helix, and a clamp that closes over the DNA–RNA hybrid during transcription. Pol I contains a Pol-II-like core consisting of five homologous (A190, A135, AC40, AC19 and A12.2) and five common (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) subunits, and the heterodimeric subcomplexes A14–A43 and A49–A34.5 (Extended Data Table 1).

To determine the Pol I structure, we previously established large-scale purification of the endogenous enzyme from the yeast *S. cerevisiae*⁹. We obtained crystals of Pol I in 2005, but these diffracted only to low resolution (C. D. Kuhn and P.C., unpublished observations). We therefore studied Pol I by electron microscopy, homology modelling and X-ray analysis of its A14–A43 subcomplex⁹. We showed that the A49–A34.5 subcomplex contains an amino-terminal dimerization domain and a carboxy-terminal tandem winged-helix (tWH) domain in A49, and solved the crystal structures of both domains¹⁰. We then used crosslinking to place A14–A43 and A49–A34.5 on the surface of a Pol I core model¹¹. Here we report the crystal structure of Pol I, which reveals new polymerase elements and provides unexpected insights into transcription regulation.

Pol I structure

In a long-standing search for well-ordered Pol I crystals, we changed parameters in yeast fermentation, protein purification and crystallization, and eventually obtained a crystal form amenable for structure determination (Methods). Purified Pol I was active in DNA-templated

RNA extension and in RNA cleavage (Extended Data Fig. 1a). The structure was solved with phases from intrinsic zinc ions and refined at 2.8 Å resolution to an R_{free} factor of 21.0% (Extended Data Table 2a, b). The structure comprises a Pol I dimer containing 8,681 amino acid residues, and lacks only the mobile A49 tWH domain and several surface loops.

The structure reveals the 10-subunit Pol I core and the subcomplexes A49–A34.5 and A14–A43 on opposite sides (Fig. 1). Compared with the largest Pol II subunit Rpb1, subunit A190 contains insertions in its domains clamp, dock, cleft, pore and funnel, and lacks parts of the clamp head, foot, jaw and cleft domains (Figs 1c and 2a, c and Extended Data Figs 2 and 4a). The jaw domain contains a long insertion that we name ‘expander’. The clamp head contains a helical insertion at the position of the archaeal polymerase subunit Rpo13 (ref. 12). Compared to the Pol II subunit Rpb2, subunit A135 contains a truncated external 1 domain, an extended protrusion, and a clamp insertion (Figs 1c and 2b, d and Extended Data Figs 3 and 4b). A heterodimer of AC40 and AC19 resembles Rpb3–Rpb11 in Pol II, but AC40 contains an additional ‘toe’ domain where the archaeal subunit D contains a domain with an iron-sulphur cluster¹³ (Fig. 1c and Extended Data Fig. 5). Subunit A12.2 binds with its N-terminal zinc ribbon domain (N-ribbon) where the homologous Rpb9 domain binds Pol II (ref. 7), but its C-ribbon binds in the pore like transcription elongation factor IIS (TFIIS)¹⁴ (Fig. 3a, b and Extended Data Fig. 6a), consistent with crosslinking¹¹.

Composite active centre

Whereas Pol II contains a ‘tunable’ active site that is transiently complemented by the C-ribbon in TFIIS¹⁴, Pol I contains a ‘composite’ active site that comprises the A12.2 C-ribbon. The C-ribbon occludes the binding site for α -amanitin, explaining why Pol I is less sensitive to this toxin than Pol II (ref. 15). The A12.2 C-ribbon contains a hairpin that reaches the active site like the catalytic hairpin of TFIIS, explaining the strong RNA cleavage activity of Pol I (ref. 9). Pol III contains a corresponding C-ribbon in its subunit C11 (ref. 16). The composite active site of Pol I and III enables efficient proofreading^{9,17} and termination^{18,19}.

The A49–A34.5 subcomplex stabilizes A12.2 (Fig. 3c), explaining why it stimulates A12.2-dependent⁹ RNA cleavage¹⁰, and why it

¹Gene Center and Department of Biochemistry, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany.

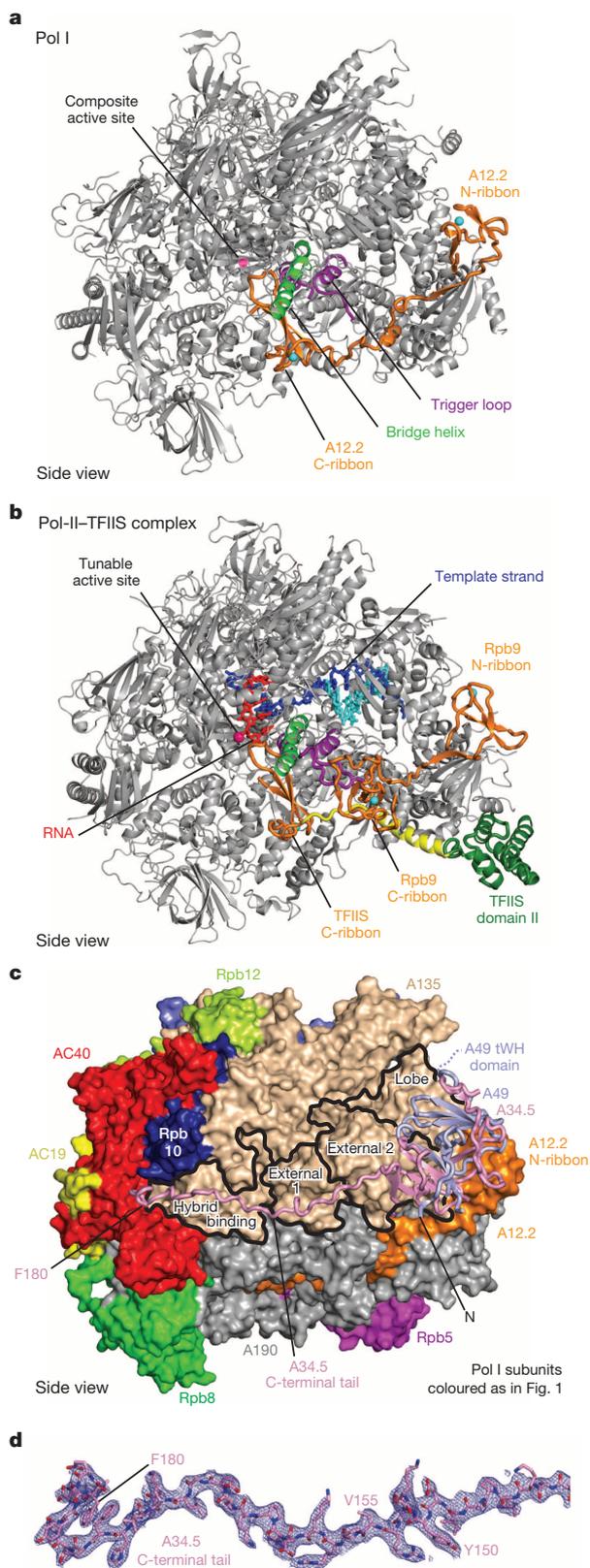


Figure 3 | Subunit A12.2 and the TFIIIF-like subcomplex A49–A34.5. **a**, Location of A12.2 (orange) within Pol I (silver) viewed from the side. The location of the N- and C-ribbon domains is indicated. **b**, Structures of Rpb9 (orange) and TFIIIS (orange, yellow, green) within Pol II¹⁴ (silver), viewed as in **a**. **c**, Ribbon model of A49–A34.5 on the Pol I core surface viewed from the side. On the basis of a comparison with the Pol II–TFIIIF complex^{22,23}, A49 and A34.5 correspond to TFIIIF subunits Tfg1 and Tfg2, respectively. The jaw-lobe module is stabilized by the A49–A34.5 subcomplex in a position close to the protrusion. **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1σ) for the positively charged A34.5 C-terminal tail that contributes to A49–A34.5 binding¹⁰.

dissociates after A12.2 deletion²⁰. A49–A34.5 stretches over the polymerase surface, to reach the AC40 toe domain with its positively charged A34.5 C-terminal tail (Fig. 3c, d and Extended Data Fig. 6b). The mobile A49 tWH domain (Fig. 1b) may close over nucleic acids in the cleft¹¹ to increase transcription processivity^{9,21}. The A49–A34.5 subcomplex is related to the Pol II initiation factor TFIIIF and to the Pol III subcomplex C37–C53, which contain similar dimerization domains located at corresponding positions^{22–24}.

Expander and cleft expansion

Compared to Pol II, the active centre cleft of Pol I is expanded by 8 Å (Fig. 4 and Supplementary Video). This results from an apparent movement of the two major polymerase modules⁷, ‘core’ and ‘shelf’, which mainly comprise subunits A135 and A190, respectively (Fig. 4a, b and Extended Data Fig. 7). Cleft expansion differs from cleft opening by swinging of the clamp⁷, and includes a relative rotation of the core and shelf modules around an axis through the active site, similar to the ratcheting observed for bacterial RNA polymerase bound by an inhibitor protein²⁵ and in a paused state²⁶. The expanded cleft is stabilized by the expander, which forms two α -helices in the active centre (Fig. 4c). The expander binds the bridge helix, which is unwound in its central region (Fig. 4d). Expander residues Asp 1385 and Asp 1388 sandwich the Pol-I-specific residue Arg 1015 in the unwound bridge helix (Fig. 4e). The expander residue Tyr 1384 inserts between the core and shelf modules, whereas Lys 1377 reaches the aspartate loop (Fig. 4e).

Cleft expansion changes the active centre that lies at the core-shelf interface (Fig. 5a). A tilting between the A190 active site domain and the A135 hybrid-binding domain alters the conformation of the catalytic aspartate loop and loop β 26– β 27 that emanate from these domains (Fig. 5b, c). In the tilted conformation, binding of catalytic metal ions and the RNA 3' end as in Pol II (refs 7, 27, 28) is impossible. Cleft contraction may rearrange the active site such that the aspartate loop can bind metal ions with its residues Asp 627, Asp 629 and Asp 631, and the A135 loop β 26– β 27 can bind the RNA 3' end with its residues Lys 916 and Lys 924. Binding of the substrate triphosphate may differ in Pol I because the A12.2 hairpin shields the A135 residues Arg 714 and Arg 957, the counterparts of triphosphate-binding Pol II residues. The A12.2 hairpin apparently interferes with closure of the triphosphate-binding trigger loop^{27,28}, but may bind the triphosphate and position metal ion B, as suggested for the TFIIIS hairpin during RNA cleavage¹⁴.

Connector and polymerase dimerization

The observed dimeric form of Pol I is established by mutual interactions of the expanded cleft of one polymerase with the stalk subcomplex A14–A43 (refs 9, 29) of the other polymerase, resulting in a handshake (Fig. 6). A14–A43 binds to the core enzyme similar to its counterparts Rpb4–Rpb7 in Pol II (ref. 30) and C17–C25 in Pol III (refs 24, 31). A43 also binds a previously unobserved N-terminal α -helix in Rpb6 (Fig. 1b), explaining the known interaction between A43 and Rpb6 (ref. 29), and loss of A43 and Rpb6 from Pol I after A14 deletion³². Stalk-mediated Pol I dimerization was also observed by electron microscopy⁶, suggesting that the dimer observed in the crystal corresponds to that formed in solution.

The flexibly linked C-terminal region of A43 folds into a ‘connector’ that invades the cleft of the adjacent polymerase (Fig. 6a, b and Extended Data Fig. 8a). The connector forms a helix (K5) that binds along the clamp coiled-coil, a β -hairpin (D1–D2) that binds the RNA exit path and the lid, and an acidic C-terminal tail that binds the polymerase switch regions 2 and 3 (Fig. 6a). The connector buries 2,350 Å² of surface area between the two polymerases, consistent with a role in dimer stabilization. The defined structure of the connector and its specific interactions with the polymerase cleft argue that connector-mediated polymerase dimerization is functionally significant. Consistent with this, the connector cannot be accommodated in a

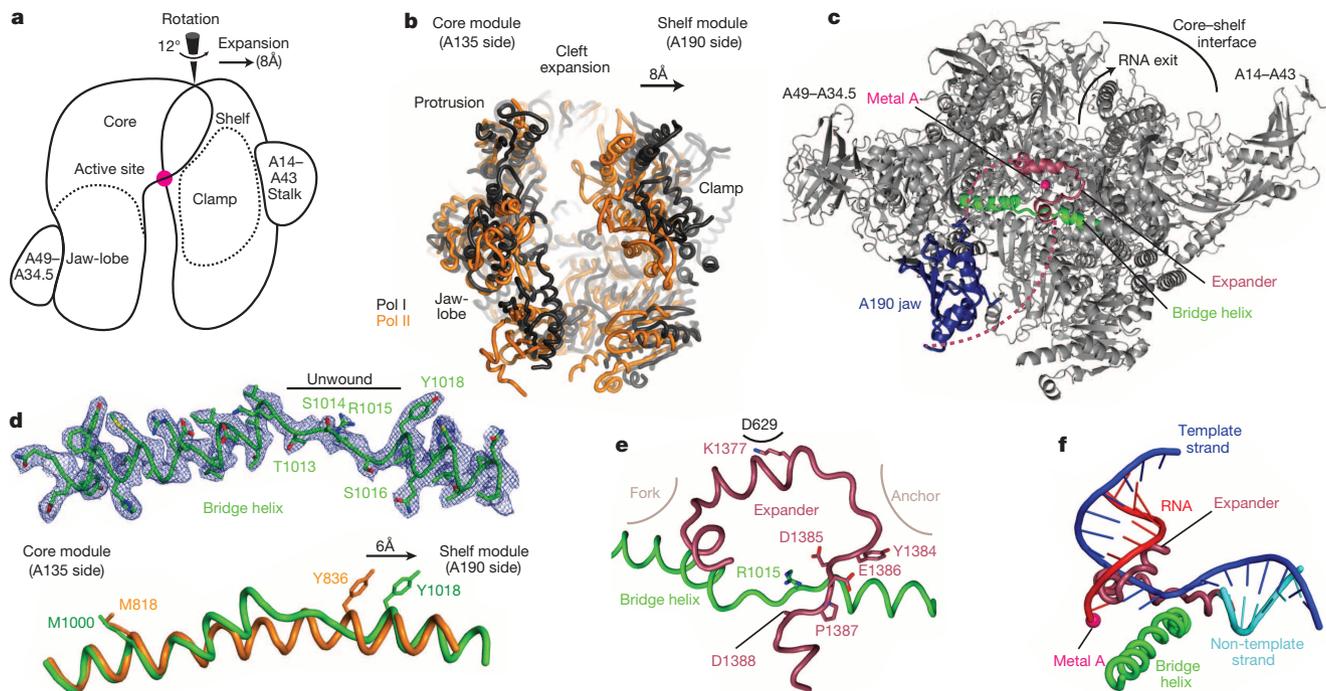


Figure 4 | Expanded cleft, expander and unwound bridge helix. **a**, Schematic of Pol I structure indicating the four mobile modules (core, shelf, clamp and jaw-lobe) and the two peripheral subcomplexes. The view is from the top⁷ as in Fig. 1, but rotated anticlockwise by 90°. The direction of cleft expansion by movement of the shelf module away from the core module is indicated by an arrow, and goes along with a 12° rotation between the two modules around the indicated axis through the active site. **b**, Expansion of the Pol I active centre cleft. Pol I (black) and Pol II (orange) were superimposed with their subunits A135 and Rpb2, respectively (left side). The view is as in **a**. **c**, A novel expander element (raspberry) that is inserted into the A190 jaw domain (blue) stabilizes the expanded Pol I cleft and a stretched and unwound bridge helix (green). **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1.3σ) for the bridge helix (green) reveals a central, unwound region. At the bottom, a comparison of the bridge helix in Pol II (orange), aligned on the core module side, reveals an expansion by 6 Å (arrow). View from the front as in Fig. 1. **e**, The expander (raspberry) stabilizes the unwound central bridge helix (green). View as in **c**. **f**, Binding of the expander and the DNA-RNA hybrid in the active centre are mutually exclusive. Superposition of the Pol I structure with the Pol II elongation complex structure (PDB code 1Y1W) reveals clashes between the expander (raspberry) and the DNA template strand (blue) and the RNA transcript (red).

View is from the top⁷, but rotated by 30° towards the front view. **d**, $2F_o - F_c$ electron density (blue mesh, contoured at 1.3σ) for the bridge helix (green) reveals a central, unwound region. At the bottom, a comparison of the bridge helix in Pol II (orange), aligned on the core module side, reveals an expansion by 6 Å (arrow). View from the front as in Fig. 1. **e**, The expander (raspberry) stabilizes the unwound central bridge helix (green). View as in **c**. **f**, Binding of the expander and the DNA-RNA hybrid in the active centre are mutually exclusive. Superposition of the Pol I structure with the Pol II elongation complex structure (PDB code 1Y1W) reveals clashes between the expander (raspberry) and the DNA template strand (blue) and the RNA transcript (red).

modelled contracted state of Pol I, and a C-terminal truncation of A43 that deletes the connector results in thermosensitivity *in vivo*²⁹, and is lethal when combined with other Pol I mutations³³.

Transcription regulation

Pol I dimerization inhibits assembly of the transcription initiation complex because it occludes the binding sites for the Pol I initiation

factors Rrn3 (refs 29, 34) and Rrn7 (refs 34–36) (Fig. 6c). Pol I dimers and monomers are in a concentration-dependent equilibrium *in vitro* (Extended Data Fig. 1b), but also monomers are predicted to be inactive because the expander overlaps with the binding sites for the DNA template strand and RNA (Fig. 4f). The expander binds the cleft of Pol I in solution, as shown by extensive crosslinking¹¹ (Extended Data Fig. 8d). The expander interferes with DNA loading in the active

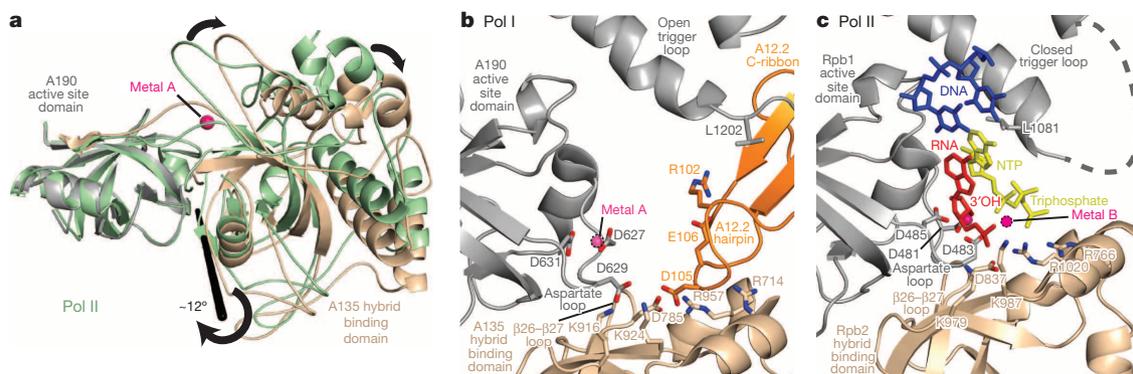


Figure 5 | Composite active site and A12.2 hairpin. **a**, Tilting between the A190 active site domain and the A135 hybrid binding domain. The active site domains of Pol I and Pol II were superposed, and the axis of rotation is indicated. The view is approximately from the side⁷. **b**, View of the Pol I active centre from the A135 side towards the clamp. The A12.2 hairpin (orange) forms an integral part of the active site. The backbone carbonyl of Asp 627 is flipped compared to its counterpart Asp 481 in Pol II, and the side chain of

Asp 629 is rotated and interacts with Lys 916 and Lys 924 in A135. To catalyse nucleotide addition, the aspartate loop must re-establish a metal-binding conformation and move closer to the A135 loop $\beta 26$ – $\beta 27$, which may contribute to binding of metal ion B. **c**, For comparison, the active site in Pol II with DNA template (blue), RNA (red), and nucleoside triphosphate substrate analogue (yellow) and the closed trigger loop are shown²⁷.

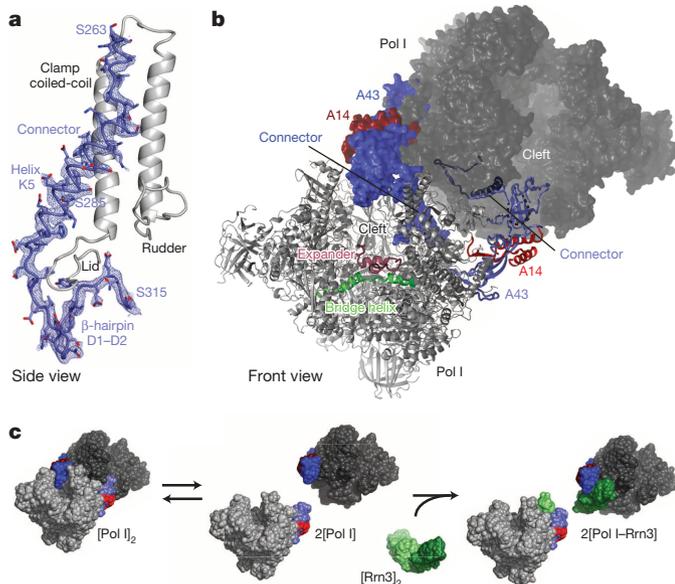


Figure 6 | Connector and Pol I dimerization. **a**, Structure of the A43 connector and corresponding $2F_o - F_c$ electron density (blue mesh, contoured at 1σ) and its binding to the coiled-coil at the inner side of the clamp (silver). **b**, Structure of the Pol I dimer and its stabilization by the A14–A43 subcomplex. One polymerase is shown as a cartoon model, the other in surface representation. The connector (blue) invades the cleft of a neighbouring polymerase. **c**, Model for Pol I initiation regulation. The binding site for the initiation factor Rrn3 (green) is occluded in the inactive Pol I dimer (silver; A14 and A43 highlighted in red and blue, respectively). Rrn3 is released from inactive Rrn3 dimers³⁴, resulting in active Pol I–Rrn3 complexes. For details compare text.

centre, because Pol I shows only weak *de novo* RNA synthesis from a tailed DNA template, in contrast to Pol II, which is much more active (Extended Data Fig. 1c). Consistent with a regulatory role of the expander, its deletion does not lead to a growth defect *in vivo* (not shown).

Pol I initiation thus requires (1) release of the connector, to generate monomers and enable Rrn3 and Rrn7 binding, (2) release of the expander, to enable DNA loading, and (3) cleft contraction, to induce a catalytically competent conformation of the active centre and enable RNA synthesis. The resulting mature DNA–RNA hybrid excludes the expander from the active centre during elongation, because Pol I is as active as Pol II when provided with a mature hybrid (Extended Data Fig. 1a).

This model for regulated Pol I initiation agrees with known functional data. First, Pol I dimers from yeast extracts are inactive in transcription^{37,38}. Second, when the connector is unavailable for Pol I dimerization, owing to a fusion with Rrn3, Pol I regulation is impaired *in vivo*³. Third, A43 phosphorylation apparently counteracts Pol I dimerization and promotes Rrn3 binding, because (1) A43 binds Rrn3 (ref. 29) and is phosphorylated in Pol I–Rrn3 complexes³⁹, (2) A43 phosphorylation sites⁴⁰ at Ser 208, Ser 220, Ser 262/263 and Ser 285 are exposed in our structure and face Rrn3 (ref. 34), and (3) Pol I dephosphorylation prevents Rrn3 binding and results in low transcription activity³⁹, and extracts from non-growing cells lack active Pol I–Rrn3 complexes³⁸. The same regulatory mechanism apparently exists in multicellular eukaryotes, because the expander is well conserved, and the connector is partially conserved (Extended Data Fig. 8b, c). Consistent with this, inactive human Pol I may be dimeric, whereas human Pol I complexes containing the Rrn3 homologue transcription initiation factor-IA (TIF-IA) are active⁴¹. Like Rrn3, human TIF-IA binds A43 (ref. 42), which may be dephosphorylated during the cell cycle by the Cdc14 phosphatase that inhibits Pol I (ref. 43).

Conclusions

The Pol I structure enables a detailed structure–function analysis of rRNA production, the first step in ribosome biogenesis. It also elucidates the evolution of related cellular RNA polymerases. Pol I and III apparently acquired a composite active site for intrinsic RNA proof-reading, 3' processing, and chain termination, to enable accurate, high-level production of stable RNA products. The composite active site probably impairs extensive RNA backtracking because the binding site for backtracked RNA⁴⁴ is occupied by A12.2. Further backtracking may occur during termination and enable cleft expansion and hybrid release, as suggested for the bacterial enzyme⁴⁵ and for Pol III (ref. 46). By contrast, the tunable active site of Pol II allows for accommodation of backtracked RNA in the pore, Pol-II-specific regulation during the elongation phase, and 3' RNA processing on the polymerase surface by the machinery for pre-mRNA cleavage and polyadenylation.

The most intriguing finding from this work, however, relates to a possible general mechanism of transcription regulation. We observe an expanded, inhibited conformation of a eukaryotic RNA polymerase that resembles a conformation of bacterial RNA polymerase associated with inactive states^{25,26}. Because the polymerase core is highly conserved, all cellular RNA polymerases can probably interconvert between an expanded inactive state and a contracted active state. State conversion will change the catalytic properties of the enzyme, because the active site is formed at the interface between the mobile core and shelf modules. Only the contracted state contains the active site in a catalytically competent conformation and binds the hybrid tightly. The conversion between polymerase states may be regulated allosterically by factors that bind at the interface of the core and shelf modules. Factors may bind in the pore, as observed for TFIIS¹⁴ and Gfh1 (ref. 25), or near the RNA exit channel, as observed for Rrn3 (refs 29, 34), the Pol II initiation factor TFIIB that stimulates RNA chain initiation allosterically⁴⁷, and the Pol II coactivator Mediator⁴⁸. The bacterial regulator catabolite activator protein⁴⁹ and the growth regulator ppGpp also bind at the RNA exit channel, and the latter was suggested to influence polymerase activity by modulating the core–shelf interface⁵⁰.

METHODS SUMMARY

A *S. cerevisiae* strain with a C-terminal affinity-purification tag on A190 was grown in a 200l fermenter. Endogenous Pol I was prepared as described⁹ with several modifications (Methods). Crystals were grown by vapour diffusion at 20 °C using 10% PEG 4000, 300 mM ammonium acetate, 50 mM HEPES, pH 7.5, and 5 mM TCEP as precipitant, and improved by microseeding in 8% PEG 4000, 300 mM ammonium acetate, 50 mM imidazole, pH 7.0, and 5 mM TCEP. Diffraction data were collected under cryo-conditions at Swiss Light Source beamlines PX1 and PX2, and at European Synchrotron Radiation Facility beamlines ID23-1 and ID29 using Pilatus 6M detectors. The structure was solved by multiwavelength anomalous diffraction from 14 intrinsic zinc ions in the Pol I dimer. Solvent flattening, non-crystallographic symmetry averaging, and phase extension to 2.8 Å resolution led to a detailed electron density map that allowed building and refinement of an atomic model.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 June; accepted 24 September 2013.

Published online 23 October 2013.

1. Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
2. Miller, O. L. Jr & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
3. Laferté, A. *et al.* The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev.* **20**, 2030–2040 (2006).
4. Drygin, D., Rice, W. G. & Grummt, I. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.* **50**, 131–156 (2010).

5. Schultz, P., Celia, H., Riva, M., Sentenac, A. & Oudet, P. Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *EMBO J.* **12**, 2601–2607 (1993).
6. Bischler, N. *et al.* Localization of the yeast RNA polymerase I-specific subunits. *EMBO J.* **21**, 4136–4144 (2002).
7. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
8. Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431–1437 (2012).
9. Kuhn, C. D. *et al.* Functional architecture of RNA polymerase I. *Cell* **131**, 1260–1272 (2007).
10. Geiger, S. R. *et al.* RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol. Cell* **39**, 583–594 (2010).
11. Jennebach, S., Herzog, F., Aebbersold, R. & Cramer, P. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.* **40**, 5591–5601 (2012).
12. Korkhin, Y. *et al.* Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7**, e1000102 (2009).
13. Hirata, A. *et al.* Archaeal RNA polymerase subunits E and F are not required for transcription *in vitro*, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol. Microbiol.* **70**, 623–633 (2008).
14. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II–TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
15. Buhler, J. M., Sentenac, A. & Fromageot, P. Isolation, structure, and general properties of yeast ribonucleic acid polymerase A (or I). *J. Biol. Chem.* **249**, 5963–5970 (1974).
16. Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D. & Cramer, P. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J. Biol. Chem.* **286**, 18701–18707 (2011).
17. Alic, N. *et al.* Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc. Natl Acad. Sci. USA* **104**, 10400–10405 (2007).
18. Prescott, E. M. *et al.* Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc. Natl Acad. Sci. USA* **101**, 6068–6073 (2004).
19. Arimbasseri, A. G., Rijal, K. & Marais, R. J. Transcription termination by the eukaryotic RNA polymerase III. *Biochim. Biophys. Acta* **1829**, 318–330 (2013).
20. Van Mullem, V., Landrieux, E., Vandenhoute, J. & Thuriaux, P. Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Mol. Microbiol.* **43**, 1105–1113 (2002).
21. Beckouët, F. *et al.* Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol. Cell Biol.* **28**, 1596–1605 (2008).
22. Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
23. Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
24. Fernández-Tornero, C. *et al.* Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J.* **29**, 3762–3772 (2010).
25. Tagami, S. *et al.* Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468**, 978–982 (2010).
26. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–441 (2013).
27. Cheung, A. C., Sainsbury, S. & Cramer, P. Structural basis of initial RNA polymerase II transcription. *EMBO J.* **30**, 4755–4763 (2011).
28. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
29. Peyroche, G. *et al.* The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J.* **19**, 5473–5482 (2000).
30. Armache, K. J., Mitterweger, S., Meinhardt, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134 (2005).
31. Jasiak, A. J., Armache, K. J., Martens, B., Jansen, R. P. & Cramer, P. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol. Cell* **23**, 71–81 (2006).
32. Smid, A., Riva, M., Bouet, F., Sentenac, A. & Carles, C. The association of three subunits with yeast RNA polymerase is stabilized by A14. *J. Biol. Chem.* **270**, 13534–13540 (1995).
33. Beckouët, F., Mariotte-Labarre, S., Peyroche, G., Nogi, Y. & Thuriaux, P. Rpa43 and its partners in the yeast RNA polymerase I transcription complex. *FEBS Lett.* **585**, 3355–3359 (2011).
34. Blattner, C. *et al.* Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev.* **25**, 2093–2105 (2011).
35. Knutson, B. A. & Hahn, S. Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. *Science* **333**, 1637–1640 (2011).
36. Naidu, S., Friedrich, J. K., Russell, J. & Zomerdijs, J. C. TAF1B is a TFIIB-like component of the basal transcription machinery for RNA polymerase I. *Science* **333**, 1640–1642 (2011).
37. Milkereit, P., Schultz, P. & Tschochner, H. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biol. Chem.* **378**, 1433–1443 (1997).
38. Milkereit, P. & Tschochner, H. A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J.* **17**, 3692–3703 (1998).
39. Fath, S. *et al.* Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. *Proc. Natl Acad. Sci. USA* **98**, 14334–14339 (2001).
40. Gerber, J. *et al.* Site specific phosphorylation of yeast RNA polymerase I. *Nucleic Acids Res.* **36**, 793–802 (2008).
41. Miller, G. *et al.* hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *EMBO J.* **20**, 1373–1382 (2001).
42. Yuan, X., Zhao, J., Zentgraf, H., Hoffmann-Rohrer, U. & Grummt, I. Multiple interactions between RNA polymerase I, TIF-1A and TAF_I subunits regulate preinitiation complex assembly at the ribosomal gene promoter. *EMBO Rep.* **3**, 1082–1087 (2002).
43. Clemente-Blanco, A. *et al.* Cdc14 inhibits transcription by RNA polymerase I during anaphase. *Nature* **458**, 219–222 (2009).
44. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
45. Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. An allosteric path to transcription termination. *Mol. Cell* **28**, 991–1001 (2007).
46. Nielsen, S., Yuzenkova, Y. & Zenkin, N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**, 1577–1580 (2013).
47. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* **493**, 437–440 (2013).
48. Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C. & Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription *in vivo*. *Science* **331**, 1451–1454 (2011).
49. Lawson, C. L. *et al.* Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14**, 10–20 (2004).
50. Zuo, Y., Wang, Y. & Steitz, T. A. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell* **50**, 430–436 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Kuhn, who obtained the first Pol I crystals. We thank C. Bäjén, S. Benkert, S. Geiger, S. Jennebach, T. Gubby and K. Maier. We thank the crystallization facility (Conti Department) and the Jentsch Department of the Max-Planck-Institut für Biochemie. Part of this work was performed at the European Synchrotron Radiation Facility at Grenoble, France, and at the Swiss Light Source at the Paul-Scherrer-Institut, Villigen, Switzerland. C.E. was supported by a PhD student fellowship of the Boehringer Ingelheim Fonds, the Elite Network Bavaria program ‘Protein Dynamics in Health and Disease’, and the Graduate Research Academy ‘RNA Biology’ of SFB960. S.S. was supported by a postdoctoral fellowship of the Alexander-von-Humboldt Foundation. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB646, TR5, GraKo1721, SFB960, CIPSM, NIM), an Advanced Grant of the European Research Council, the Jung-Stiftung, and the Vallee Foundation.

Author Contributions C.E. planned and carried out experiments and crystal structure determination. S.S. advised on experimental and crystallographic work. A.C.C. performed computational crystallographic analysis. D.K. contributed to computational crystallography and model building. P.C. designed and supervised research and prepared the manuscript, with contributions from all authors.

Author Information Atomic coordinates and structure factors of the Pol I crystal structure have been deposited with the Protein Data Bank under accession number 4C2M. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (cramer@LMB.uni-muenchen.de).

METHODS

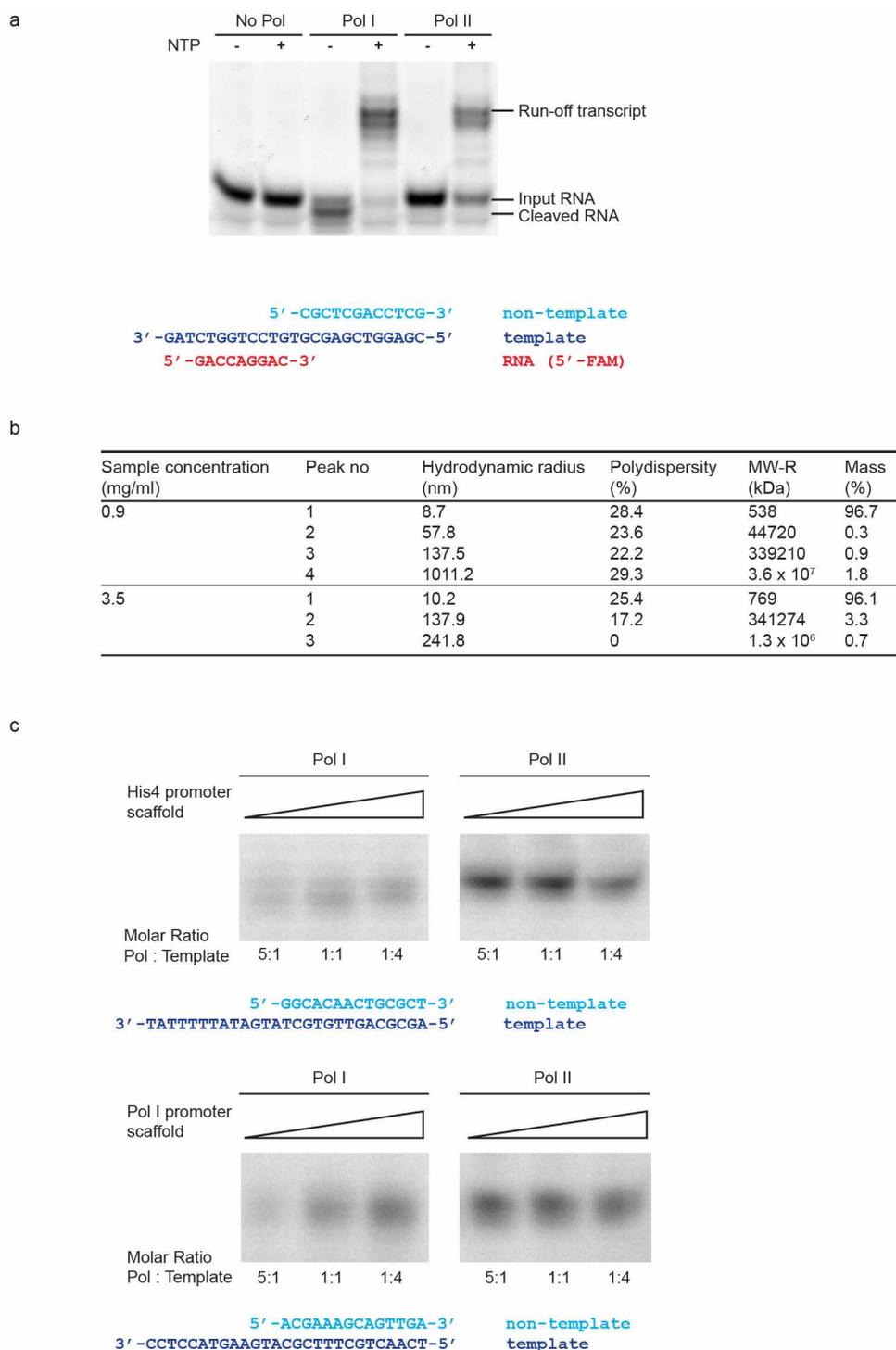
Pol I preparation. Pol I enzyme used for crystallization was essentially obtained as described⁹, using a protocol that was based on a procedure originally established by the Tschochner laboratory⁵¹, but with several changes. Pol I was purified from the expression-optimized strain CB010 expressing a C-terminal Flag/10×histidine-tagged subunit A190. Cells were grown to an attenuation (*D*) at 600 nm of 2.0 in YPD medium in a 200 l fermenter and collected using a continuous-flow centrifuge. Cells were suspended and flash-frozen in buffer A (150 mM HEPES, pH 7.8, 60 mM MgCl₂, 20% (v/v) glycerol, 5 mM dithiothreitol (DTT), 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM benzamidine, 60 μM leupeptine and 200 μM pepstatine). For purification, 500 g of yeast cells were lysed by bead beating in buffer A after the addition of ammonium sulphate to 400 mM at 4 °C. The lysate was cleared by centrifugation (30 min at 10,000g) and ultracentrifugation (90 min at 30,000g) and dialysed overnight against buffer B (50 mM potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10% (v/v) glycerol, 10 mM β-mercaptoethanol, 1 mM PMSF and 1 mM benzamidine). After centrifugation (60 min at 18,500g) the pellet was re-suspended in buffer C (1.5 M potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10 mM imidazole, 10% (v/v) glycerol, 10 mM β-mercaptoethanol, 1 mM PMSF and 1 mM benzamidine) and incubated with 8 ml Ni-NTA beads (Qiagen) for 4 h. Bound protein was washed with five column volumes (CV) buffer C, followed by 5 CV buffer D (300 mM potassium acetate, 20 mM HEPES, pH 7.8, 1 mM MgCl₂, 25 mM imidazole, 10% (v/v) glycerol, 10 mM β-mercaptoethanol) and eluted in 5 CV buffer E (buffer D but containing 200 mM imidazole). The eluate was loaded on a 10/100 MonoQ anion exchange column (GE Healthcare) and eluted with a gradient from 0.3 to 2.0 M potassium acetate in buffer F (20 mM HEPES, pH 7.8, 1 mM MgCl₂, 10% (v/v) glycerol and 5 mM DTT). Eluted Pol I was diluted from 1.1 to 0.2 M potassium acetate, loaded on a 5/50 MonoS cation exchange column (GE Healthcare) and eluted with a gradient from 0.2 to 1.0 M potassium acetate in buffer F including a plateau at 350 mM potassium acetate. Pol I eluted at 490 mM, was concentrated to 500 μl and applied to a Superose 6 10/300 size exclusion column (GE Healthcare) in buffer G (60 mM ammonium sulphate, 5 mM HEPES, pH 7.8, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM DTT). A single peak contained up to 1.8 mg of homogenous and pure Pol I.

Pol I crystallization. Freshly purified Pol I was subjected to commercial screens and to crystallization conditions that had previously produced RNA polymerase crystals from different organisms. MRC-2 microplates (96-well format; SWISSCI) were set up with 100 nl sitting drops using the Phoenix pipetting device (Art Robbins). Initial crystals were obtained in 10% PEG 4000, 300 mM ammonium acetate, 50 mM HEPES, pH 7.5, and 5 mM TCEP at 20 °C. These crystals formed thin plates that poorly diffracted X-rays. Optimization and micro-seeding yielded brick-like crystals with a size of up to 300 × 60 × 40 μm using as reservoir solution 8% PEG 4000, 300 mM ammonium acetate, 50 mM imidazole, pH 7.0, and 5 mM TCEP. After a two-step transfer to the same buffer containing an additional 20% PEG400, crystals were flash-frozen and stored in liquid nitrogen.

Pol I crystal structure determination. Diffraction data were collected at the Swiss Light Source in Villigen, Switzerland, on beamlines PX1 and PX2 as well as at the European Synchrotron Radiation Facility in Grenoble, France, on beamlines ID23-1 and ID29 using Pilatus 6M detectors. Data were processed with XDS⁵² and showed P1 symmetry (Extended Data Table 2a, b). A two-fold non-crystallographic symmetry (NCS) axis was detected by calculation of a self-rotation function. Molecular

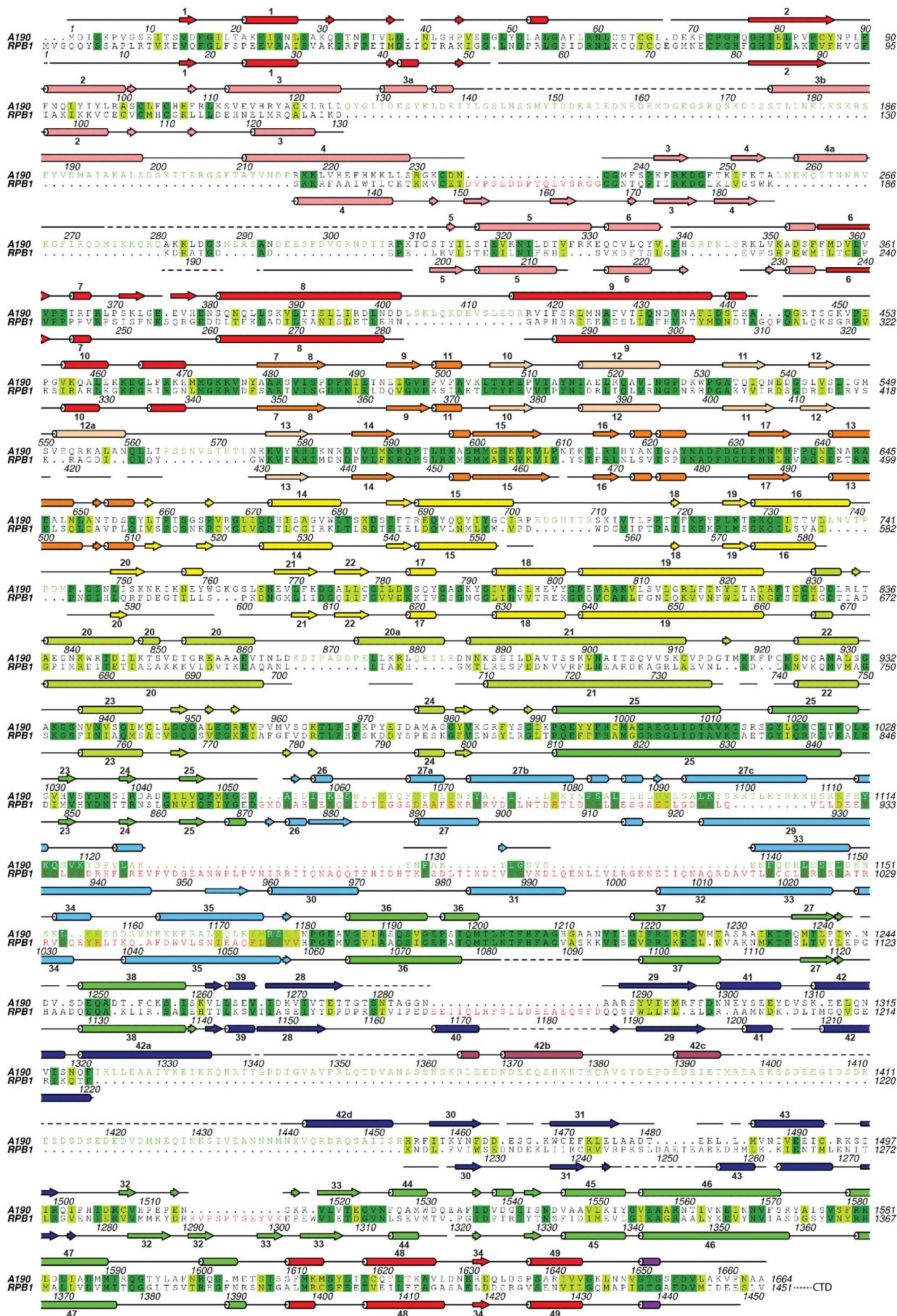
replacement using as search models Pol II, a Pol I homology model¹¹ or a Pol I electron microscopic envelope⁹ failed. We could, however, use anomalous diffraction from intrinsically bound zinc ions for phasing. After performing a fluorescence scan, three data sets were collected from a single crystal at the peak, inflection, and low-energy remote wavelengths (Extended Data Table 2b). From the peak data set, the positions of 14 zinc ions (7 in each Pol I complex) were determined with SHELX C and D⁵³. The occupancies calculated by SHELX were between 100% and 58% for the first 14 sites, followed by a drop to ~25% for additional peaks that were not considered. Zinc positions were refined using autoSHARP⁵⁴ in a multiwavelength anomalous diffraction experiment using four wavelengths (Extended Data Table 2a, b). Two Pol II-based homology models of the Pol I core were fitted to match the experimental zinc positions using the program MOLOC. Initial phases at 3.5 Å resolution were obtained with autoSHARP followed by solvent flattening with SOLOMON⁵⁴ using the positioned homology models as an initial mask. The program PARROT⁵⁵ performed automated NCS averaging and extended phases to 2.8 Å resolution. An atomic model of Pol I was built with COOT⁵⁶ starting from the positioned homology models. The PHENIX suite⁵⁷ was used to refine the structure with manually assigned rigid body groups considering domain boundaries, and independent TLS parameterization for a total of 24 TLS groups. NCS restraints with groups defined according to the rigid bodies and individual atomic displacement parameters were applied during refinement. In the later stages of refinement, prominent water molecules and two bound sulphate ions were placed and validated by visual inspection. The final model has an *R*_{free} factor of 21.0% at 2.8 Å resolution and excellent stereochemistry (Extended Data Table 2a). For visualization, secondary structure was assigned with DSSP⁵⁸ and by inspection.

51. Tschochner, H. A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc. Natl Acad. Sci. USA* **93**, 12914–12919 (1996).
52. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
53. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
54. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
55. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
58. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
59. Bischler, N. *et al.* Specific interaction and two-dimensional crystallization of histidine tagged yeast RNA polymerase I on nickel-chelating lipids. *Biophys. J.* **74**, 1522–1532 (1998).
60. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
61. Geiger, S. R., Kuhn, C. D., Leidig, C., Renkawitz, J. & Cramer, P. Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Crystallogr. F* **64**, 413–418 (2008).
62. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
63. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).



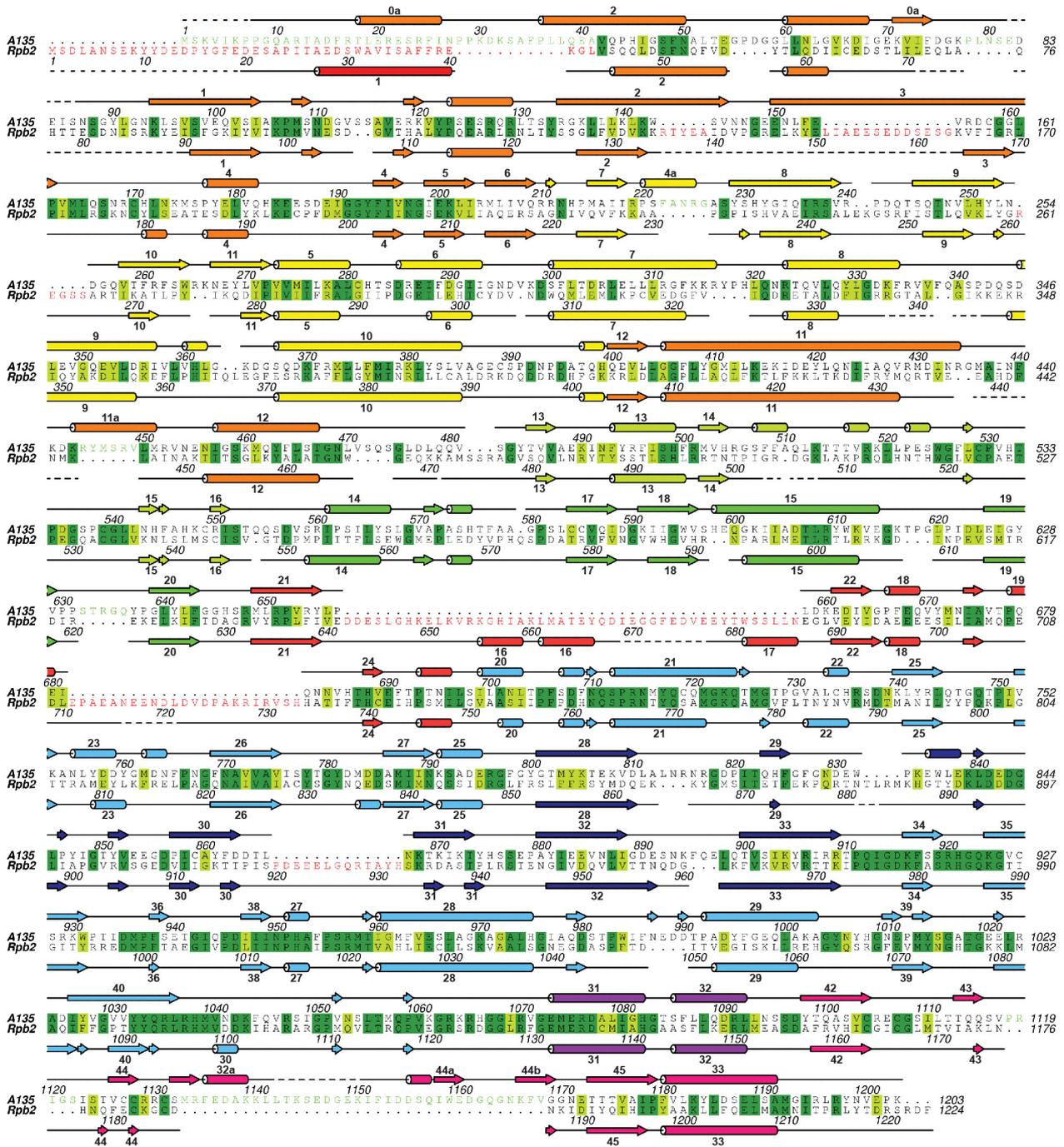
Extended Data Figure 1 | Activity of purified Pol I *in vitro*. **a**, Pol I is active in RNA extension and cleavage. Purified *S. cerevisiae* Pol I and Pol II can extend RNA in a DNA–RNA scaffold in the presence of nucleoside triphosphate substrates (NTPs). In the absence of nucleoside triphosphate substrates, Pol I cleaves RNA. Transcription assays were performed as described⁹ using the DNA–RNA scaffold shown. After running a 20% acrylamide UREA gel, RNA was detected by fluorescence. **b**, Dynamic light scattering is consistent with a concentration-dependent Pol I dimer–monomer equilibrium in solution. Note that the sample with the higher Pol I concentration shows an increased

hydrodynamic radius and apparent (estimated) molecular mass for the predominant peak by mass that accounts for over 96%. The method cannot distinguish monomeric and dimeric species. Thus, the estimated molecular mass observed at 3.5 mg ml^{-1} Pol I concentration (769 kDa) arises from a mixture of monomers (590 kDa) and dimers (1,180 kDa). **c**, *De novo* RNA synthesis activity on tailed DNA template. Assays were performed as described⁴⁷ using the DNA template shown. After running a 20% acrylamide UREA gel, *de novo* synthesized, radioactive RNA was detected by phosphoimaging.



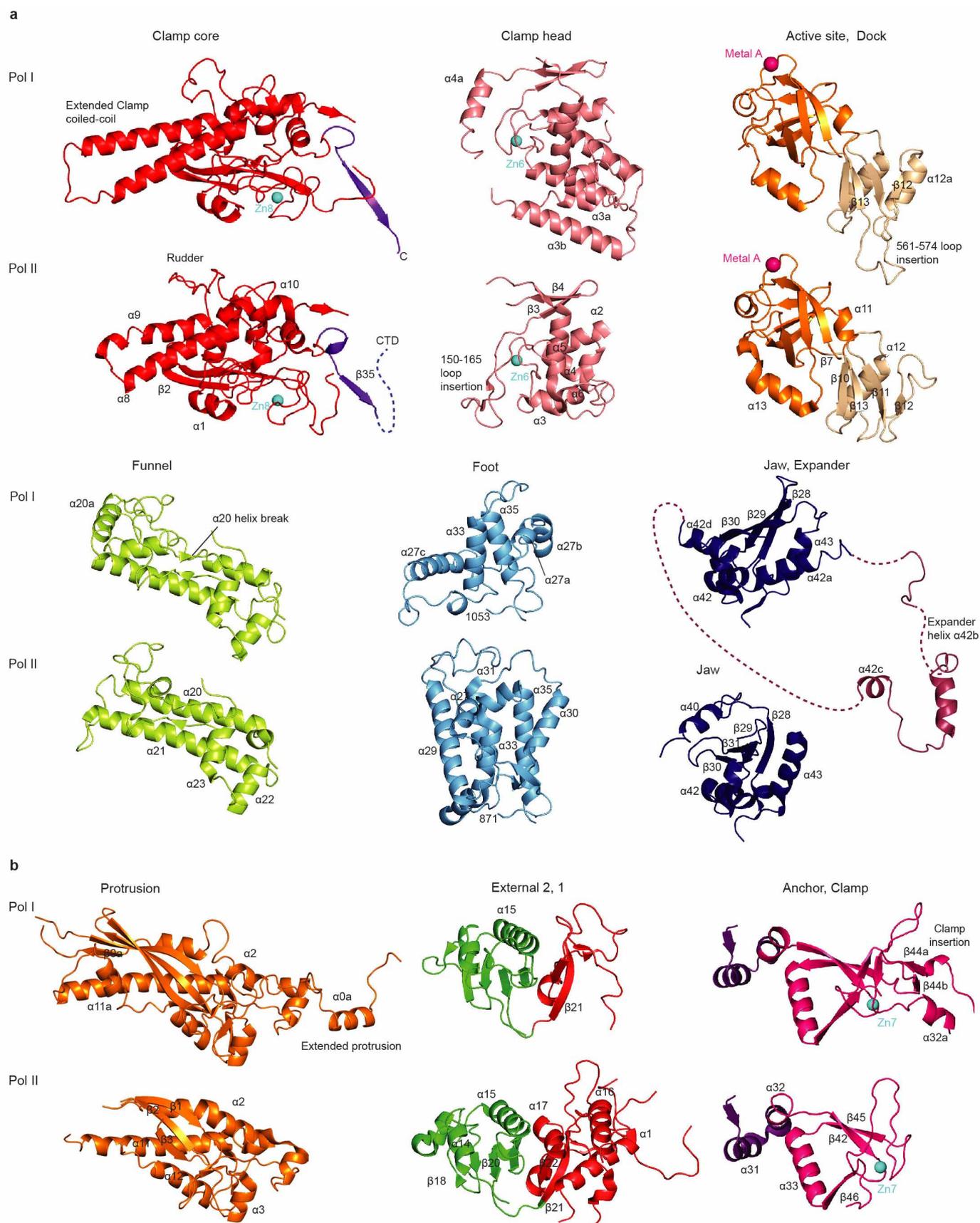
Extended Data Figure 2 | Structure-based alignment of A190 and Rpb1.
Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the

alignment for A190 and Rpb1, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).



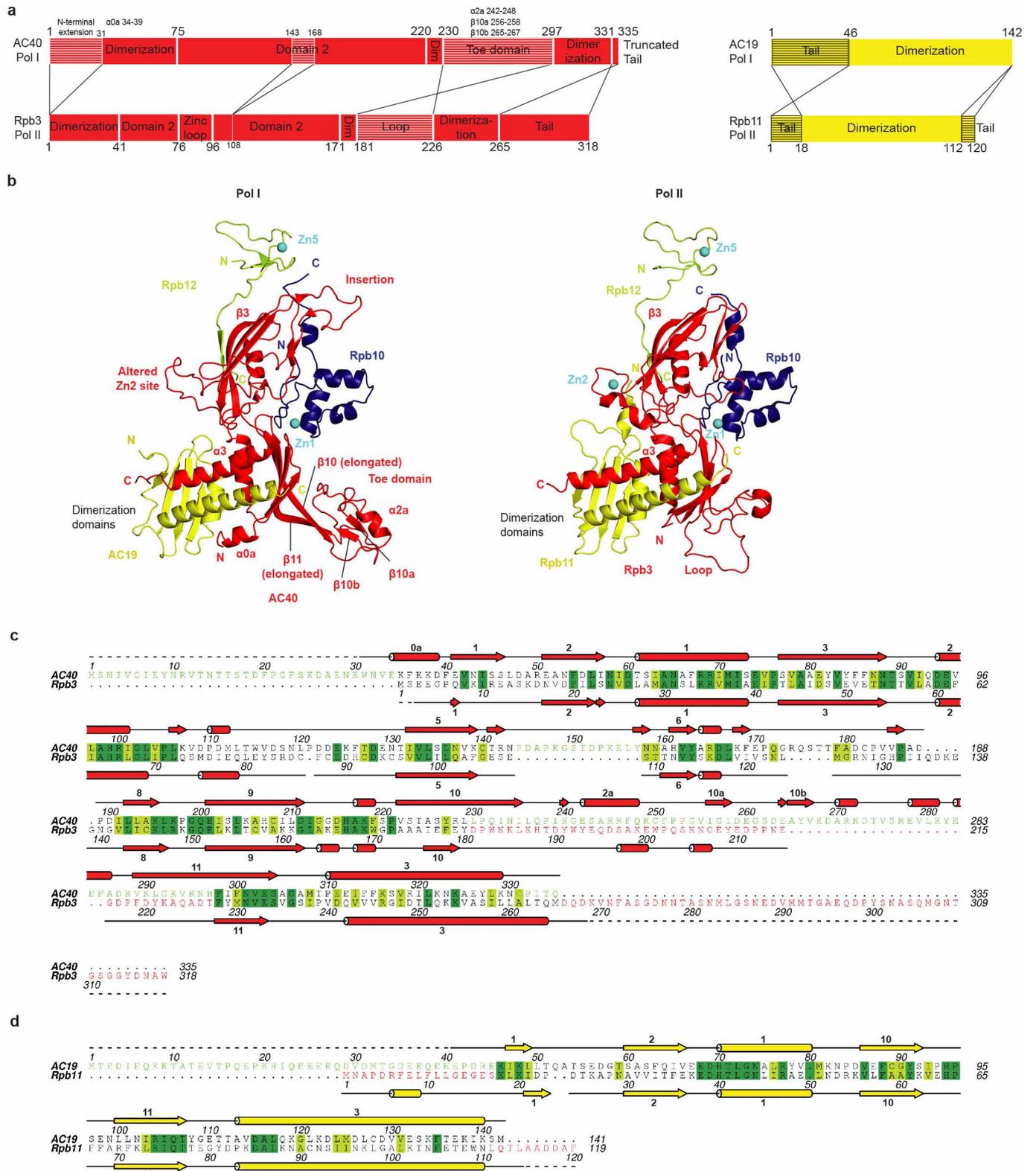
Extended Data Figure 3 | Structure-based alignment of A135 and Rpb2. Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the

alignment for A135 and Rpb2, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).



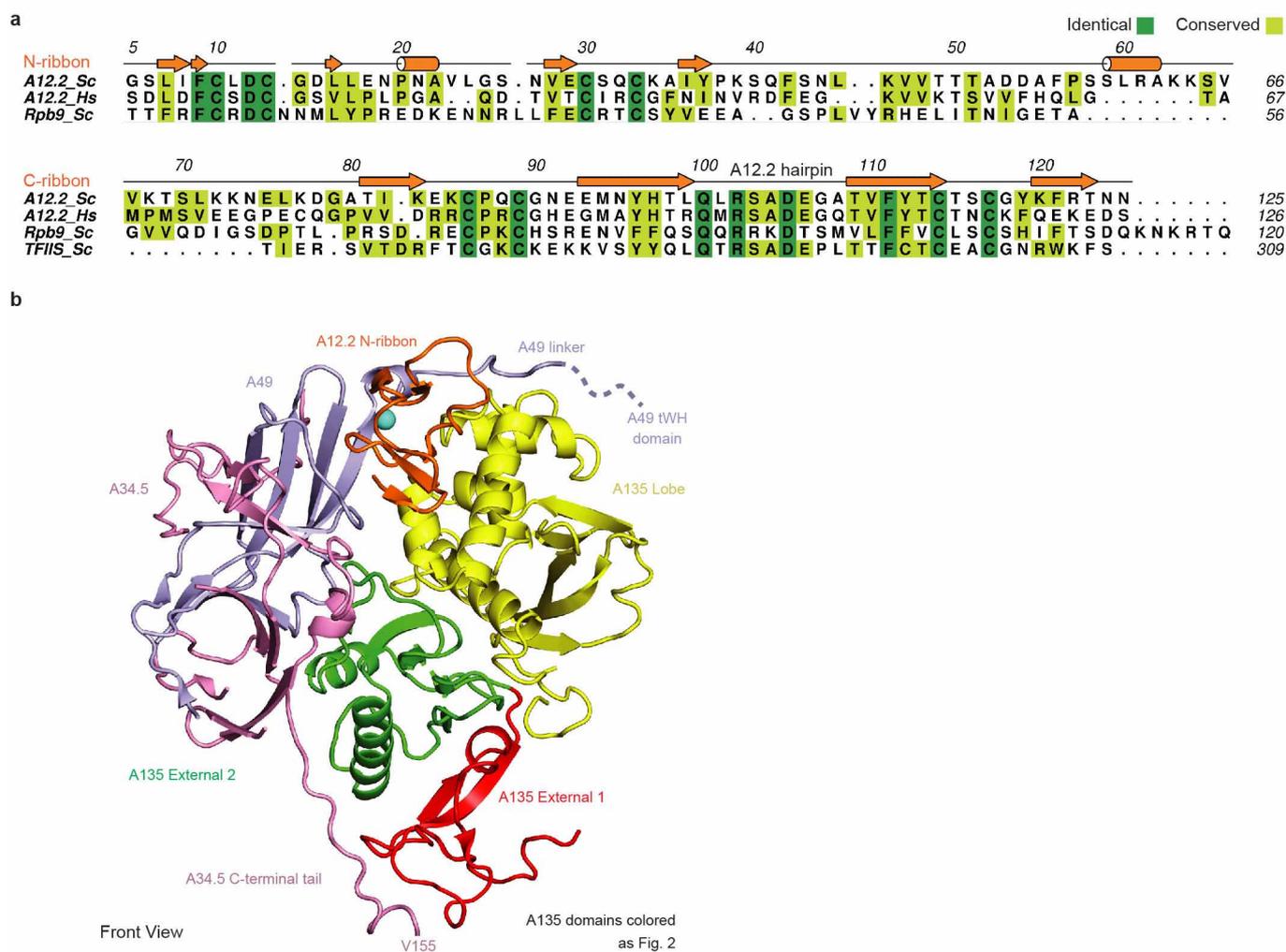
Extended Data Figure 4 | Detailed comparison of A190–A135 with Rpb1–Rpb2. a, Comparison of A190 domain structures (top) that differ significantly from their corresponding Pol II regions (bottom). Labelling of corresponding secondary structure elements is as for Pol II (ref. 7). New or lacking secondary

structure elements are labelled. New elements were named according to the preceding Pol II element with small letters added alphabetically for subsequent elements. **b,** Comparison of A135 domain structures (top) that differ significantly from their corresponding Pol II regions (bottom) as in Extended Data Fig. 3.



Extended Data Figure 5 | Structure of the subassembly AC40–AC19–Rpb10–Rpb12. **a**, Schematic of domain structures. AC40 and AC19 were aligned with homologous Pol II subunits Rpb3 and Rpb11, respectively. Organization and annotation as in Fig. 2a. **b**, Ribbon view from the ‘back’ of the enzyme⁷ (left, Pol I; right, Pol II). **c**, **d**, Structure-based alignments of AC40 and

Rpb3 (c) and of AC19 and Rpb11 (d). Invariant and conserved residues are highlighted in green and light green, respectively. Secondary structure elements are indicated above and below the alignment for AC40 and Rpb3, respectively (cylinders, helices; arrows, strands). Residues that form different folds in Pol I or form Pol-II-specific folds are in green or red, respectively (compare Fig. 1c).

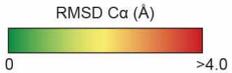


Extended Data Figure 6 | A12.2 and A49–A34.5. **a**, Subunit A12.2 amino acid sequence alignment. The sequence of yeast *S. cerevisiae* A12.2 (Sc) was aligned with that of the human subunit (*Homo sapiens*; Hs), and the N- and C-terminal domains were aligned with their counterparts in the Pol II subunit

Rpb9 and the Pol II elongation factor TFIS¹⁴, respectively. **b**, Interaction of A49–A34.5 with Pol I core domains. The view is from the front of the enzyme⁷ (Fig. 1). Different Pol I subunit domains are coloured as in Figs 2 and 3.

Domain superpositions between Pol I and pseudo-Pol II (Pol I domains organised into a Pol II arrangement)

RMSD C α (Å)	Active site	Anchor	Dock	Rpb5	Foot	Cleft	Rpb5	Clamp Core	Clamp A135	Clamp Head	A43	A14	A12.2 N-rib	Jaw Upper	Lobe	Ext2	Ext1	Rpb8	Pore1	Funnel	Fork	Hybrid Binding	AC40	Rpb10	AC19	Protrusion	Rpb12	Wall
Active site	0.0	1.3	1.3	1.1	1.2	1.0	1.8	2.6	3.1	5.4	8.2	8.5	60.5	19.4	7.7	21.1	7.3	2.4	1.5	2.1	3.2	4.1	3.4	3.6	3.8	4.8	6.2	6.3
Anchor	0.5	0.0	1.8	1.0	1.1	0.9	1.0	0.6	2.1	5.2	4.5	4.7	65.8	12.5	5.5	24.6	9.6	1.7	2.6	3.3	5.1	6.3	5.3	5.7	6.2	5.5	7.8	6.9
Dock	0.7	2.0	0.0	1.2	0.8	1.8	3.3	3.9	4.1	5.2	7.7	7.7	51.9	21.3	9.8	31.3	13.3	5.2	4.5	6.2	8.1	9.0	7.7	8.0	7.7	4.6	8.2	9.4
Rpb5	1.4	0.9	1.7	0.0	1.4	0.7	1.7	1.7	3.3	3.3	5.0	5.2	78.6	10.1	10.2	36.2	15.4	6.2	5.6	5.8	8.2	10.0	9.0	8.4	10.6	5.2	12.3	12.6
Foot	1.7	1.8	1.5	2.0	0.0	1.6	1.7	2.3	4.6	7.4	10.5	10.9	56.1	14.5	15.2	32.0	11.5	4.0	2.8	2.5	3.4	4.6	4.4	4.1	8.7	6.5	12.4	14.4
Cleft	1.8	1.6	2.3	1.3	1.3	0.0	0.8	3.2	6.4	7.4	12.2	12.2	40.5	10.3	8.8	19.6	7.2	6.7	6.0	6.6	7.9	8.5	7.3	9.2	8.5	9.1	8.6	9.6
Rpb5	3.5	1.9	2.5	1.7	2.3	1.4	0.0	3.5	7.8	7.1	14.6	14.6	33.6	17.2	18.5	27.8	11.2	11.4	9.5	9.2	11.4	12.8	11.6	13.1	13.8	11.6	12.3	11.3
Clamp_Core	2.3	1.1	2.1	2.4	1.9	2.9	1.6	0.0	1.9	3.9	2.9	2.9	34.5	27.9	18.4	31.6	16.6	6.5	8.6	11.1	13.4	14.6	12.5	14.4	13.2	10.3	8.8	6.7
Clamp_A135	4.0	2.5	3.0	3.5	3.0	3.7	2.4	1.4	0.0	4.1	2.4	2.4	9.6	21.6	15.8	36.7	18.2	9.0	10.0	11.9	14.5	15.9	14.0	15.4	14.5	10.4	10.8	8.8
Clamp_Head	2.5	2.7	3.9	3.1	2.1	3.6	2.5	4.1	7.8	0.0	10.4	10.2	42.0	25.1	20.5	32.4	16.5	9.5	10.1	11.7	14.3	16.2	14.1	16.0	15.9	11.3	11.5	7.4
A43	10.6	7.8	8.7	9.1	9.3	8.8	6.1	5.2	2.7	10.1	0.0	0.8	56.4	11.4	13.8	47.6	23.7	19.1	17.4	18.1	21.2	23.0	21.2	22.1	21.9	14.5	15.4	14.9
A14	11.1	8.5	9.2	10.2	9.6	10.0	7.3	6.0	1.8	11.7	0.8	0.0	19.7	16.1	14.9	46.7	23.4	18.9	17.7	18.9	21.8	23.4	21.4	22.9	21.7	15.8	14.6	14.2
A12.2 N-Rib	14.2	20.7	11.9	18.4	14.4	19.8	18.5	14.6	13.5	14.8	17.7	5.7	0.0	19.9	14.1	19.4	26.8	17.6	27.8	16.6	31.1	18.7	17.1	32.0	10.9	21.8	13.5	18.0
Jaw_Upper	9.8	12.0	12.8	11.5	9.8	12.5	12.5	14.7	20.6	9.0	26.7	26.1	44.2	0.0	8.1	8.5	8.8	11.4	12.5	16.9	16.8	16.6	14.1	18.7	13.0	15.5	14.1	15.9
Lobe	14.0	16.7	17.7	17.0	13.3	16.8	16.3	18.7	22.8	19.8	26.0	25.3	28.1	14.6	0.0	6.1	6.4	7.1	8.0	11.9	10.4	9.9	8.5	11.8	6.0	10.0	15.1	18.1
External2	15.7	17.9	18.5	18.6	14.0	17.0	16.3	19.4	21.1	22.3	25.5	24.9	64.0	18.3	4.1	0.0	2.9	12.5	7.7	4.8	4.5	4.9	5.4	4.5	7.8	5.9	14.1	13.3
External1	11.0	12.8	13.4	13.6	9.2	12.2	11.5	14.5	16.4	18.5	22.5	22.2	78.6	23.0	8.0	3.7	0.0	9.5	4.9	2.0	2.5	2.9	2.8	2.7	5.2	3.5	11.0	10.8
Rpb8	5.4	8.0	6.8	7.0	4.0	5.5	6.4	10.2	9.3	11.5	20.7	21.4	54.1	36.3	23.5	29.3	10.9	0.0	1.2	3.6	3.3	3.2	2.9	2.9	4.3	5.6	12.3	13.6
Pore1	3.2	5.0	4.2	4.4	2.3	3.5	4.3	6.7	6.5	8.7	14.6	15.1	70.2	24.5	13.9	22.6	7.1	1.2	0.0	1.4	1.2	1.8	2.0	1.2	2.9	5.2	8.8	9.6
Funnel	8.7	11.0	11.4	10.7	7.5	9.5	9.6	13.3	15.2	14.2	23.4	23.4	88.0	19.9	12.3	12.9	3.3	5.8	2.8	0.0	1.4	2.1	2.2	2.4	1.9	3.6	13.2	12.5
Fork	8.6	9.4	10.7	10.4	7.8	9.6	8.5	10.0	11.8	14.5	15.1	14.8	75.8	23.1	8.1	9.2	3.6	7.4	3.6	1.7	0.0	1.0	1.2	1.4	3.7	1.4	9.3	8.7
Hybrid_Binding	5.4	6.6	6.9	7.1	4.6	6.4	6.1	8.2	9.9	11.9	16.8	16.8	27.7	30.9	15.0	11.5	3.5	5.8	2.6	1.4	0.7	0.0	0.9	1.0	1.7	3.7	7.4	6.9
AC40	5.4	8.3	7.0	7.8	4.5	7.4	8.6	11.3	13.0	12.4	22.5	22.6	56.4	44.7	24.9	12.3	4.7	5.7	1.9	1.9	1.8	1.0	0.0	1.5	1.9	5.6	6.5	5.5
Rpb10	5.1	6.9	6.8	7.9	3.4	7.4	7.1	9.3	12.1	13.8	20.3	20.3	74.2	42.3	22.1	6.0	3.7	7.1	2.6	1.0	0.9	0.6	0.9	0.0	2.5	3.4	3.3	5.7
AC19	6.6	9.4	7.4	8.5	6.0	8.1	9.8	11.8	11.5	11.2	18.9	19.2	60.2	36.0	19.4	18.7	4.8	3.5	2.7	3.6	2.9	1.7	1.1	2.3	0.0	8.1	3.5	5.5
Protrusion	9.9	10.5	11.7	12.2	8.6	10.8	8.9	10.4	11.0	17.3	11.8	11.4	78.5	37.2	18.2	16.2	8.2	9.2	3.7	1.7	2.8	3.9	2.6	3.5	6.2	0.0	5.9	5.3
Rpb12	6.0	4.4	5.8	6.6	5.5	5.8	3.1	3.4	5.8	13.7	12.2	12.2	83.7	45.6	23.8	16.4	5.6	10.5	4.8	2.9	2.7	2.4	2.5	3.1	3.8	1.9	0.0	3.1
Wall	7.2	6.1	7.5	7.8	6.7	6.8	4.5	4.7	4.9	13.2	8.8	8.8	85.7	38.9	19.4	18.3	8.6	9.6	4.9	3.7	3.7	3.5	3.2	4.4	2.9	3.2	2.9	0.0
Module	Shelf			Clamp			Stalk		Jaw-Lobe			Core																



Extended Data Figure 7 | Shift in domain positions between Pol I and Pol II. Individual domain fragments of Pol I were isolated and superposed onto their counterparts in Pol II and subsequently recombined to form a ‘pseudo-Pol II’ model. This pseudo-Pol II model was then superposed onto the complete Pol I structure over a single common domain (indicated by column headings).

The resulting root mean squared deviation (r.m.s.d.) C α (Å) value for every individual domain (indicated by row headings) was calculated using PyMol and coloured according to value (green to yellow = 0.0 to 3.0 Å, orange = 3.0 to 4.0 Å, and red >4.0 Å).

Extended Data Figure 8 | Structure and conservation of the connector and expander.

a, The connector binds the adjacent polymerase. The electron density is not continuous between the end of the A43 N-terminal part (residue Ile 250) and the beginning of the A43 connector in the adjacent polymerase. The connector cannot be assigned to A43 from the same polymerase because the observed distance between the last residue in the stalk and the first residue in the connector (>72 Å) cannot be spanned with only 12 residues. By contrast, the distance from Ile 250 to the connector in the cleft of the neighbouring polymerase is 18 Å and easily spanned by the missing residues. We note that dimeric forms of Pol I were previously observed in two-dimensional arrays⁵⁹. **b**, Amino acid sequence alignment of the connector region in *S. cerevisiae* (*Sc*); *Candida glabrata* (*Cg*); *Homo sapiens* (*Hs*) and *Mus musculus* (*Mm*). Secondary structure elements are indicated above the alignment (K5 helix, cylinder;

strands D1, D2; arrows). Residues that are involved in the interface with the Pol I clamp and cleft and showed a buried surface area in excess of 40 Å² are indicated with an asterisk above the alignment. Buried surfaces were calculated with the PISA server⁶⁰. A structure-based alignment of the A14–A43 stalk residues was published^{9,61} and is not included here. **c**, Amino acid sequence alignment of the expander region in *S. cerevisiae*, *C. glabrata*, *H. sapiens* and *M. musculus*. Secondary structure elements are indicated above the alignment (helices, cylinders; strands, arrows). Residues that are involved in the interface with the Pol I cleft and showed a buried surface area in excess of 40 Å² are indicated with an asterisk above the alignment. Buried surfaces were calculated with the PISA server⁶⁰. **d**, Previously obtained crosslinks¹¹ of the expander element map to the Pol I cleft. Crosslinks to A190 and A135 are indicated in grey and wheat respectively.

Extended Data Table 1 | RNA polymerase subunits in *S. cerevisiae*

RNA polymerase	Pol I	Pol II	Pol III [*]
Core enzyme subunits	A190	Rpb1 (22.2%) [†]	C160
	A135	Rpb2 (25.9%) [†]	C128
	AC40	Rpb3 (18.9%) [†]	AC40
	AC19	Rpb11 (20.0%) [†]	AC19
	A12.2 [‡]	Rpb9 [‡] (18.3%) [†]	C11 [‡]
	Rpb5	Rpb5	Rpb5
	Rpb6	Rpb6	Rpb6
	Rpb8	Rpb8	Rpb8
	Rpb10	Rpb10	Rpb10
	Rpb12	Rpb12	Rpb12
Stalk subcomplex A14/A43	A14	Rpb4 (1.8%) [†]	C17
	A43	Rpb7 (12.3%) [†]	C25
TFIIF-like subcomplex	A49	Tfg1 (TFIIF α)	C37
	A34.5	Tfg2 (TFIIF β)	C53

* Pol III also contains the trimeric subcomplex C82–C34–C31 that was not included here.

[†] Percentage amino acid residues of the Pol II subunit that are identical in the corresponding Pol I subunit, based on the structure-based alignments in Extended Data Figs 2, 3 and 5 and published data⁹.

[‡] The C-terminal zinc-ribbon domain of A12.2 and C11 corresponds to domain III of the Pol II elongation and transcript cleavage stimulatory factor TFIIS.

Extended Data Table 2 | Diffraction data and statistics

a

Crystal I-IV	
Data collection*	
Space group	P1
Cell dimensions	
a, b, c (Å)	122.7, 139.0, 209.6
α , β , γ (°)	108.1, 95.4, 93.8
Resolution (Å)	40-2.8 (2.87-2.80) [†]
R_{sym}	24.0 (201.8)
$I/\sigma I$	9.61 (1.52)
Completeness (%)	99.9 (99.8)
Redundancy	13.1 (11.5)
$CC_{(1/2)}^{\ddagger}$ (%)	99.6 (54.1)
Refinement	
Resolution (Å)	40-2.80
No. of reflections	320,814
$R_{\text{work}}/R_{\text{free}}$	16.8/21.0
No. amino acid residues	8,681
No. of atoms	
Protein	68,820
Ligand/ion	24
Water	214
B-factors (Å ²)	
Protein	81.1
Ligand/ion	130.4
Water	53.8
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.26
Ramachandran [§]	
Preferred/allowed/disallowed (%)	94.5/4.6/0.9

b

Crystal V	Peak	Inflection	Remote
Data collection*			
Wavelength (Å)	1.28288	1.28382	1.29639
Cell dimensions			
a, b, c (Å)	122.0, 139.2, 209.1		
α , β , γ (°)	108.3, 95.1, 94.2		
Resolution (Å)	40-3.5 (3.59-3.50) [†]	40-3.5 (3.59-3.50)	40-3.5 (3.59-3.50)
R_{sym}	15.1 (54.3)	17.0 (74.1)	21.3 (99.8)
$I/\sigma I$	16.54 (6.0)	8.14 (2.0)	6.53 (1.4)
Completeness (%)	98.0 (97.9)	97.2 (95.5)	97.1 (95.5)
Redundancy	9.8 (9.9)	5.0 (5.0)	5.0 (4.9)
$CC_{(1/2)}^{\ddagger}$ (%)	99.7 (94.5)	99.0 (67.9)	98.5 (47.3)

a, Diffraction data and refinement statistics. **b**, Multiwavelength anomalous diffraction data collection statistics.

* Diffraction data were collected on a Pilatus 6M detector (**a**) at beamline X10SA of the Swiss Light Source and ID23-1 and ID29 of the European Synchrotron Radiation Facility, and (**b**) at beamline X06SA of the Swiss Light Source. Diffraction data were processed with XDS². In **a**, data were collected at 0.25° or 0.1° increments (180° or 360° in total, respectively) at a wavelength of 1.0 Å. Data sets of four crystals were merged. The R_{pim} values were 0.072 (overall, 40–2.8 Å), 0.026 (40–15.3 Å) and 0.659 (2.85–2.80 Å). In **b**, data were collected in 0.1° increments (1080° in total). As a high-energy remote data set, the data given in **a** were used.

[†] Numbers in parentheses refer to the highest resolution shell.

[‡] $CC_{1/2}$ = percentage of correlation between intensities from random half-data sets⁶².

[§] Calculated with MolProbity⁶³.