8-20-2024

Structural Basis for Substrate Binding and Selection by Human Mitochondrial RNA Polymerase

Karl Herbine

Ashok Nayak

Dmitry Temiakov

Article

Structural basis for substrate binding and selection by human mitochondrial RNA polymerase

Karl Herbine [1], Ashok R. Nayak [1] & Dmitry Temiakov [1] ✉

The mechanism by which RNAP selects cognate substrates and discriminates between deoxy and ribonucleotides is of fundamental importance to the fidelity of transcription. Here, we present cryo-EM structures of human mitochondrial transcription elongation complexes that reveal substrate ATP bound in Entry and Insertion Sites. In the Entry Site, the substrate binds along the O helix of the fingers domain of mtRNAP but does not interact with the templating DNA base. Interactions between RNAP and the triphosphate moiety of the NTP in the Entry Site ensure discrimination against nucleosides and their diphosphate and monophosphate derivatives but not against non-cognate rNTPs and dNTPs. Closing of the fingers domain over the catalytic site results in delivery of both the templating DNA base and the substrate into the Insertion Site and recruitment of the catalytic magnesium ions. The cryo-EM data also reveal a conformation adopted by mtRNAP to reject a non-cognate substrate from its active site. Our findings establish a structural basis for substrate binding and suggest a unified mechanism of NTP selection for single-subunit RNAPs.

DNA-dependent RNA polymerases (RNAPs) must select cognate substrates and discriminate in favor of ribonucleotides during the nucleotide addition cycle (NAC). Historically, elucidation of this process at a structural level was first attempted using single-subunit DNA-dependent RNAPs, which are found in T7-like and N4 bacteriophages, mitochondria, and chloroplasts[1–4]. Understanding the mechanisms behind substrate incorporation in single-subunit RNAPs is important for enzymatic synthesis of RNA, efficient production of mRNA for vaccines such as the Moderna and Pfizer Covid19 vaccines[5], development of anti-cancer drugs targeting human mitochondrial RNAP (mtRNAP, or POLRMT)[6], and sensible design of nucleotide-based prodrugs to prevent their off-target effects on human mitochondrial nucleotide polymerases[7].

Two substrate binding sites, termed Pre-insertion and Insertion sites, have been identified by structural studies in bacteriophage RNAPs[1,2]. In the Pre-insertion site, the elongation complex (EC) has completed the last nucleotide addition cycle to assume a post-

translocated conformation in which the 3' end of RNA has moved to the product site, and the fingers domain (residues 954-1016) adopts an "open" conformation. The phosphate moiety of NTP is bound along the O helix of the fingers domain, while the base makes incipient base-pairing interactions with the *n* base of DNA, allowing for "pre-selection" of a cognate substrate[2]. In contrast, in the Insertion site, which has been observed in both T7 and N4 RNAPs, as well as in yeast mtRNAP, the fingers domain adopts a "closed" conformation, and the substrate NTP is fully inserted into the active site and is poised for catalysis[1,3,8]. In both Pre-insertion and Insertion sites, the 2' OH group of the ribose ring is "sensed" by a conserved tyrosine residue, allowing for the selection of ribonucleotides, which are present in cells at significantly higher concentrations than deoxyribonucleotides.

Mitochondrial RNAPs share sequence and structural homology with T7 RNAP, however, structural information on substrate-bound mtRNAP is lacking[9–12]. In this study, we use cryogenic electron microscopy (cryo-EM) and single particle analysis to obtain high-resolution

[1]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, 1020 Locust St, Philadelphia, PA 19107, USA.
✉e-mail: dmitry.temiakov@jefferson.edu

structures of substrate-bound human mtRNAP ECs to elucidate mechanisms of substrate entry, binding, and selection.

Results and discussion

### The overall structure of the substrate-bound transcription elongation complexes

We assembled functional transcription elongation complexes using human mtRNAP, transcription elongation factor TEFM, and a bubble nucleic acid scaffold containing an 8 bp RNA-DNA hybrid, in the presence of a non-hydrolyzable substrate - α,β−methylene ATP (Fig. 1a–c). In Complex ONE, the incoming ATP was complementary to the acceptor base n (TMP) in the template strand of DNA and thus served as a cognate substrate. In Complex TWO, the n base was changed to AMP, converting the incoming ATP to a non-cognate substrate. The assembled

complexes were subjected to cryo-EM followed by single particle analysis (Supplementary Fig. 1-8, Supplementary Tables 1, 2). Because of the high dynamics associated with the substrate binding process, 3D variability analysis was used to improve the resolution of substrate-bound complexes (see Materials & Methods), resulting in two major clusters (Cluster 1 and 6, Supplementary Fig. 2) and two corresponding high-resolution structures (Fig. 1d, Supplementary Fig. 1a, b) for Complex ONE. In Structures I and II, the 3′ end of the RNA is found in the product site, indicating that these complexes represent the post-translocated step of the NAC. TEFM is bound to mtRNAP atop the intercalating hairpin and interacts with the upstream and downstream DNA duplex, as demonstrated in previous studies[11] (Supplementary Fig. 3b, 4b). Clusters 2-5 exhibited conformations resembling those observed in Structures I and II, suggesting potential intermediate states along the
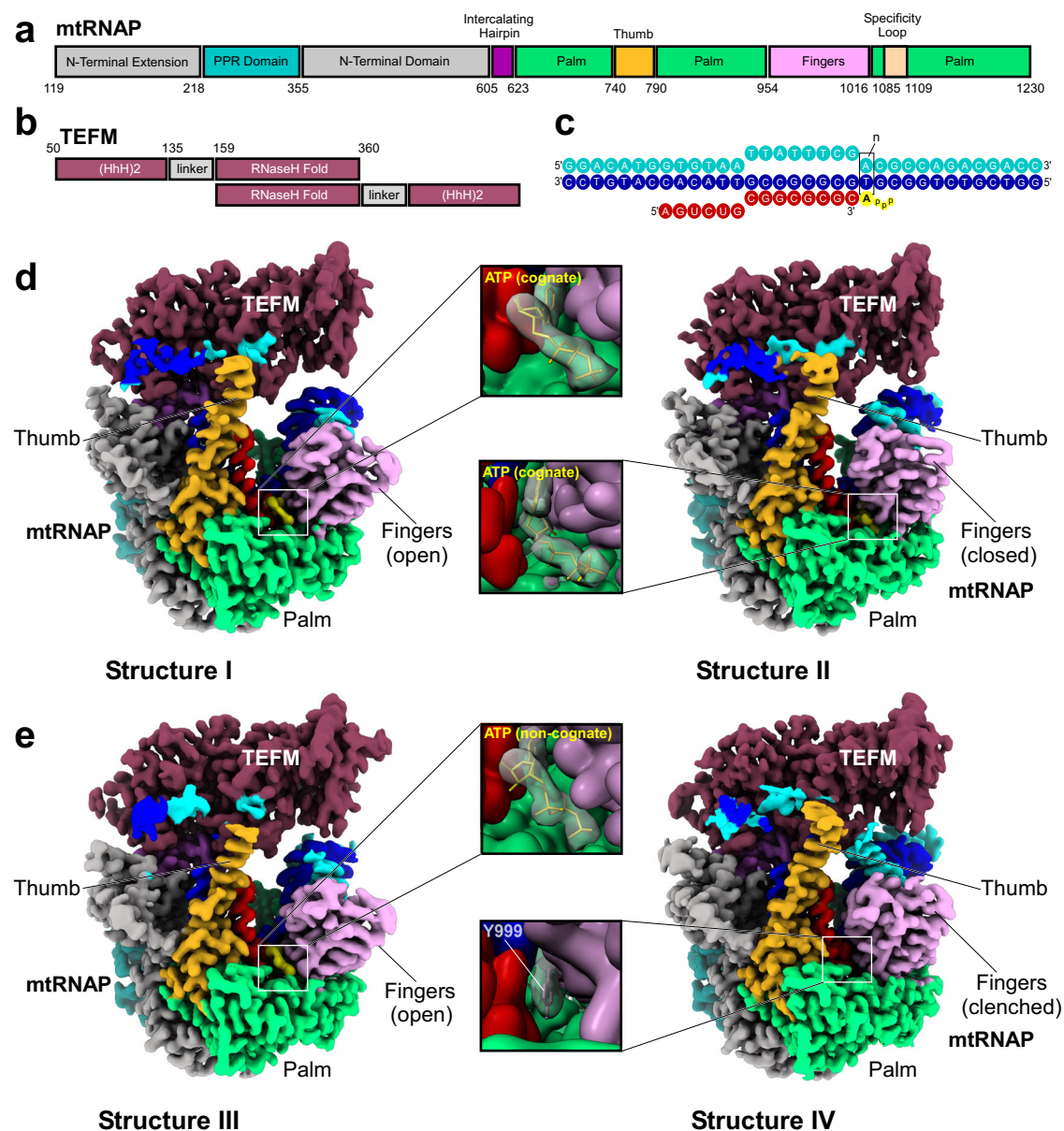
Fig. 1 | Structures of human mitochondrial transcription elongation complexes with bound substrate. a,b Schematic representation of the functional domains in mtRNAP and TEFM. The color code is used throughout. **c** Scaffold design used to assemble EC (Complex ONE). The n base pair (shown in the open box) was inverted in Complex TWO. NT DNA strand is shown in cyan, TS – in blue, and RNA – in red. The ATP analog is depicted as a yellow circle. **d** Cryo-EM densities of substrate-bound ECs from Complex ONE in two different conformational states. Close-up views of the active sites reveal cryo-EM density of the bound substrate ATP (yellow sticks). **e** Cryo-EM densities of ECs from Complex TWO in two different conformational states. Close-up views of the active sites reveal cryo-EM density of the bound non-cognate substrate ATP (Structure III) and Y999 residue (Structure IV).
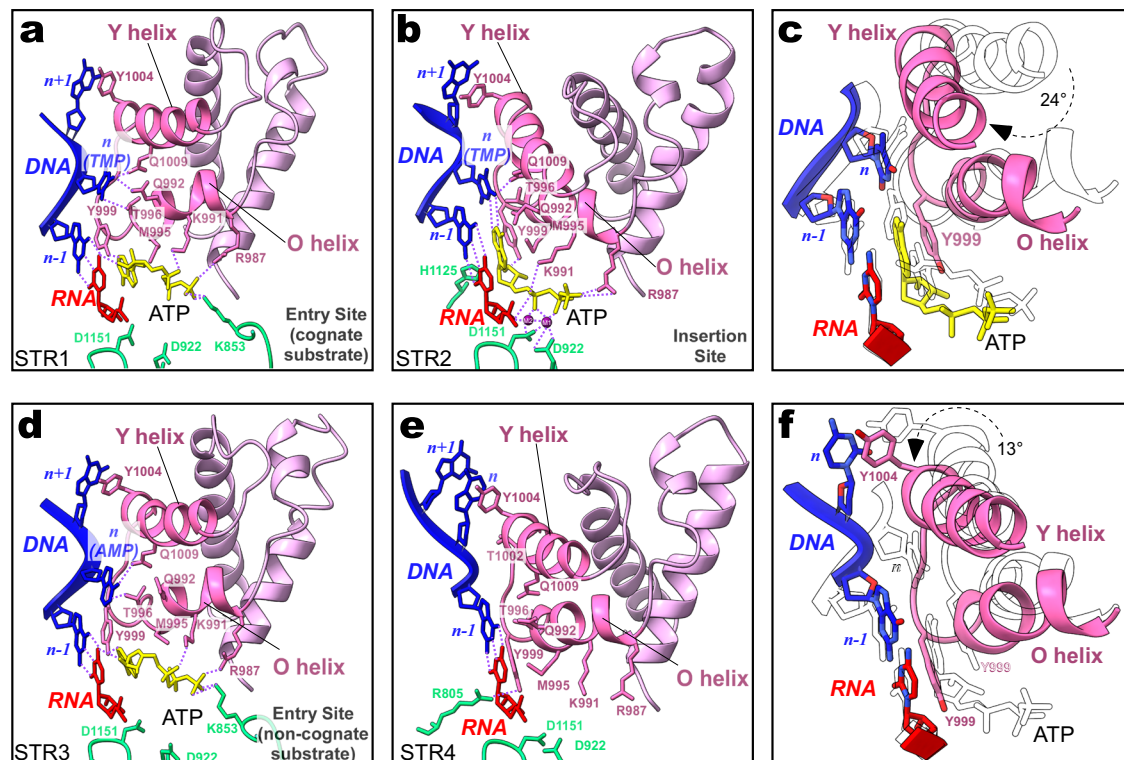
Fig. 2 | Structural basis for substrate binding and selection by human mitochondrial RNAP. a,b Close-up views of Substrate Entry and Substrate Insertion Sites of mtRNAP. The structures were aligned using the conserved residues in the palm domain. DNA, RNA, ATP, the palm domain, and the fingers domain are colored blue, red, yellow, green, and pink, respectively. Catalytic Mg^{2+} ions (M1 and M2) are shown as purple spheres; M1 was modeled based on homology modeling with T7 RNAP. Hydrogen bonds are shown as blue dashed lines. **c** Movement of the O and Y helices from Substrate Entry (white) and Substrate Insertion (colored) Sites. **d** Close-up view of the Substrate Entry Site with the bound non-cognate substrate. **e** Close-up view of the Substrate Rejection Complex. **f** Movement of the O and Y helices from closed (white) to clenched (colored) conformation.

pathway of substrate ATP binding. However, 3D refinement of these intermediates failed to produce high-resolution cryo-EM maps of mtRNAP region near its active site, likely due to the high mobility of the fingers domain (Supplementary Fig. 2).

Processing of data collected for Complex TWO (non-cognate substrate) also resulted in six major clusters, two of which represented a post-translocated EC with the bound substrate ATP (non-cognate substrate, Structure 3), and another four - the "closed" conformation of the EC without a bound substrate (Structure 4) (Fig. 1e, Supplementary Fig. 1, 5–8). Similarly to Structures I and II, the 3' end of the RNA in Structures III and IV is found in the product site of these complexes, indicative of the post-translocated state.

The phosphate moiety of the incoming ATP in Structure I and Structure III is bound along the O helix of the fingers domain, whereas the nucleotide moiety has not entered the active site of mtRNAP (Fig. 2a, d). Since no changes in protein and RNA-DNA topology have been detected as compared to the previously resolved structure of human mtRNAP EC-TEFM complex[11], we surmise that the obtained structures represent a nucleotide entry point into the active site, and therefore refer to them as Substrate Entry Complex. As will be discussed below, this is different from the Pre-insertion site previously described[2] in that there is no incipient base-pair yet formed with the acceptor *n* template DNA base.

In contrast, Structure II revealed significant changes in the RNAP active site, resulting in compaction of the fingers domain over the bound substrate, complete insertion of substrate into the active site, and recruitment of catalytic Mg^{2+} ions (Fig. 2b). This is a Substrate Insertion Complex, a conformation previously observed in single-subunit bacteriophage RNAPs[1,3]. Notably, we did not detect a bound substrate in the Insertion Site when ATP was non-complementary to

the *n* base in Complex TWO (Supplementary Fig. 1, 5–8). Instead, a significant number of particles represented a post-translocated EC characterized by a more tightly closed conformation of the fingers subdomain, termed here as "clenched," and lacked a bound substrate (Structure IV, Fig. 2e). In contrast to Structures I-III, the acceptor *n* base of template DNA in Structure IV is expelled from the active site and is found base-paired with the complementary base of the non-template strand, unavailable for substrate binding. The conformation of this complex, termed Substrate Rejection, appears to be more compact than Structure II. In this configuration, the fingers domain is pressed against the palm domain, while TEFM makes additional interactions with the downstream and upstream DNA, resulting in clear density for several additional bases of the NT strand (Supplementary Fig. 1d). The trajectory of the nucleic acids and the conformation of the fingers domain in Structure IV is similar to that observed in previous X-ray studies[10]. Overall, the Substrate Entry and Substrate Insertion Complexes provide a structural basis for initial substrate binding and selection by mitochondrial RNAPs, as described below.

Substrate entry complex

In the Entry Site, the phosphate moiety of ATP is bound by three positively charged residues: K853 (palm subdomain), K991, and R987 of the O helix (Fig. 2a). The sugar moiety of the incoming substrate is stabilized by a hydrogen bond between the oxygen in 3' OH of ATP and the 3' OH of the RNA, preventing the latter from nucleophilic attack on the substrate's phosphate moiety. The Y999 residue of the O helix is found stacking upon the *n-1* base of the DNA and is seen in a position to make a hydrogen bond with the N1 of the purine ring of ATP, while M995 is stacking upon it. The acceptor DNA base *n* is inserted between O and Y helices and is "sandwiched" between T996 and two glutamine
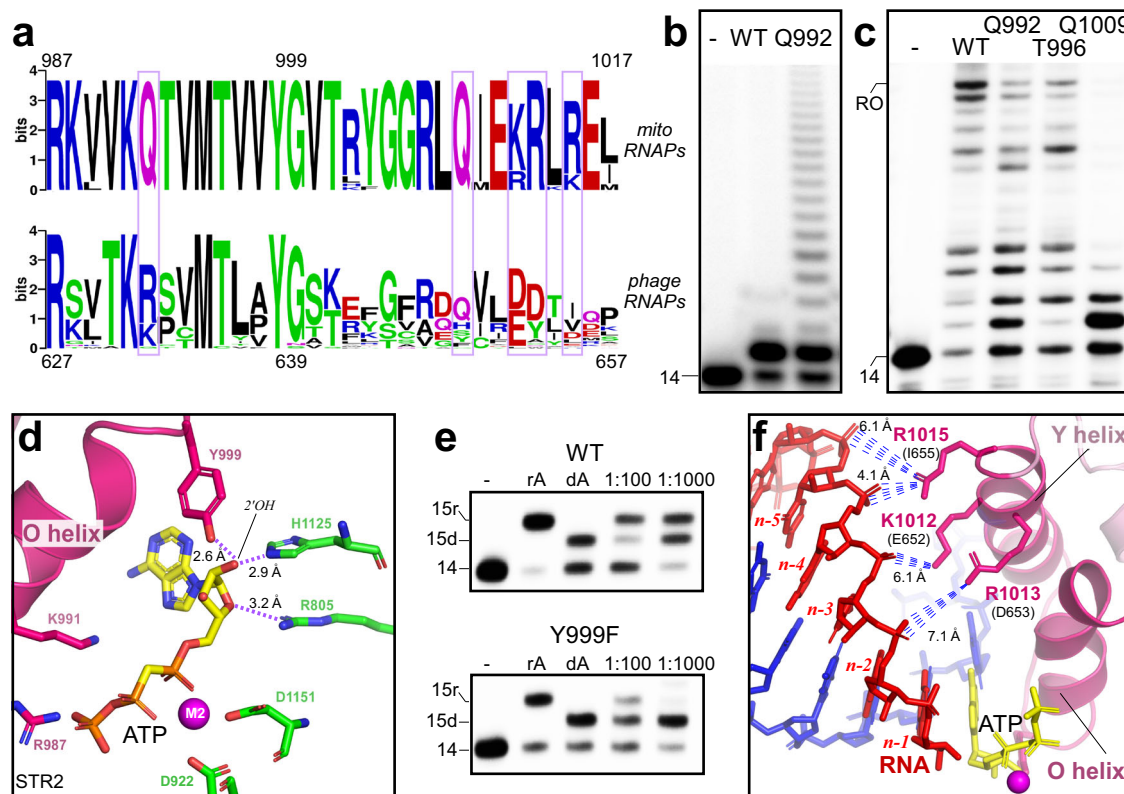
Fig. 3 | Properties of human mitochondrial RNAP variants defective in translocation and rNTP recognition. a Sequence logos of O/Y helices in mitochondrial (top) and bacteriophage (bottom) RNAPs. The residues that lack conservation between mitochondrial and phage RNAPs and are implicated in the NAC are shown in purple boxes. **b** Transcript slippage activity of the Q992A variant of mtRNAP. The size of the RNA primer (14 nt) is indicated. **c** Primer extension assay performed with mtRNAP variants Q992A, T996A, and Q1009A. **d** A close-up view of ribose moiety of NTP binding and recognition. **e** rNTP/dNTP competition assay. The RNA primer (14 nt) was extended in the presence of rNTP/dNTP mixtures, as indicated. The reaction products differ in mobility due to rAMP (15r) or dAMP (15d) incorporation. **f** Interactions between the Y helix and RNA in the Insertion Site of mtRNAP EC. Long-range electrostatic interactions are shown as dotted blue lines.

residues, Q992 and Q1009, making it inaccessible for base-pairing with the incoming substrate (Fig. 2a). This arrangement resembles sequestering of the *n* DNA base between helices O and O1 in DNA polymerase I, suggesting similarity of the translocation mechanisms in Pol A family of DNA and RNA polymerases[13].

The Q1009 residue hydrogen bonds to the N3 atom of dTMP at the *n* position (Fig. 2a). Since the acceptor *n* DNA base is not in the vicinity of the active site, and its eventual position there is occupied by Y999, no discrimination between cognate and non-cognate substrates is possible at the Entry Site, as evidenced by Structure III, which represents EC with a non-cognate substrate bound in the Entry Site (Fig. 2d, Supplementary Fig. 1c). Likewise, no residues are located in proximity to the 2′ OH group of ATP, suggesting that both ribo- and deoxy- NTPs can bind in the Entry Site. The binding of a cognate substrate has been detected in a similar site for a non-cognate NTP in T7 RNAP[14] and in the structurally related DNA polymerase I[15].

Substrate insertion complex

In Structure II, the substrate ATP is bound in the Insertion Site, poised for incorporation into the RNA chain (Fig. 2b). This site has been extensively characterized in Pol A family of RNA and DNA polymerases[1,16–19]. Delivery of ATP into the Insertion Site requires "closing" of the fingers subdomain by rotating the 5-helix bundle (res 954-1043) 24° away from the palm subdomain around an axis connecting residues V955 and V997, nearly perpendicular to the RNA-DNA heteroduplex (Fig. 2c, Supplementary Movie 1–3). This results in moving the Y999 residue away from its stacking position with the *n-1* base of the template DNA strand and placing the *n* DNA base into the active site, where it forms Watson-Crick interactions with ATP.

Notably, the *n* DNA base maintains interactions with the O helix residues Q1009 and T996, while Q992 is seen to hydrogen bond with the base of the incoming ATP (Fig. 2b). In phage RNAPs, residue Q992, which is conserved among mitochondrial RNAPs, is substituted with arginine and is not in position to interact with the substrate in the insertion site (Fig. 3a). The Q992A variant of mtRNAP is catalytically active but allows repetitive incorporation of substrate NTP ("slippage"), consistent with the role of this residue in sequestering the 3′ end of RNA in the Insertion Site (Fig. 3b). Residues T996 and Q1009 are conserved among single-subunit RNAPs (Fig. 3a); their substitutions in T7 RNAP result in variants that are significantly less processive than the wild-type enzyme[20,21], consistent with their role in translocation proposed here. MtRNAP variants T996A and Q1009A, as well as slippage-prone Q992A were notably defective in both transcription initiation (Supplementary Fig. 9a) and primer extension assays (Fig. 3c). Strikingly, the Q1009A mtRNAP was only able to extend the RNA primer by a single nucleotide (Fig. 3c).

In the Insertion Site, the phosphate moiety of ATP is still bound along the O-helix, however, some contacts are lost, and some new interactions are established, as compared to the Substrate Entry Complex. Thus, the K853 residue of the palm subdomain is no longer in position for hydrogen bonding, whereas the Y956 residue in the O-helix interacts with the β phosphate of ATP (Fig. 2b). The extended conformation of the triphosphate moiety is stabilized by the presence of two catalytic Mg^{2+} atoms, which are coordinated by D922, D1151, and G923 residues and positions the α phosphate of ATP within 3 Å from the 3′ end of the RNA. Movement of the Y999 residue from its stacking position with the *n-1* DNA base in the Entry Site to its eventual position in the Insertion Site brings its hydroxyl group within interacting

distance with the 2′ OH group of ribose, providing a structural basis for selectivity in favor of ribonucleotides (Fig. 3d). In addition, the imidazole ring of a conserved in single-subunit RNAPs H1125 residue is located within hydrogen bonding distance with the 2′ OH of ribose of substrate ATP in the Insertion Site (Fig. 3d). Indeed, the Y999F mtRNAP variant is significantly more permissive to dNTP incorporation in a substrate competition assay, when a mixture of ribo/deoxy NTP is provided (Fig. 3e). Similarly to T7 RNAP[22,23], the double substitution variant, Y999F/H1125A mtRNAP, exhibits an even higher propensity to incorporate dNTPs as compared to the Y999F mtRNAP variant (Supplementary Fig. 9c).

The most notable difference between the conserved residues of the fingers domain in phage and mitochondrial RNAPs is the presence of three positively charged amino acids in the Y helix – K1012, R1013, and R1015 in the latter group of polymerases (Fig. 3f). In contrast, in phage RNAPs, the corresponding residues bear a negative charge (Fig. 3a). The positively charged residues in the Y helix of mtRNAP are involved in long-range electrostatic interactions with the phosphate backbone of RNA, which appear to stabilize the closed conformation of the fingers domain, observed in Structure II and Structure IV.

Substrate rejection complex

Complex TWO dataset contains a different conformation of the EC represented by a large majority of the particles. In this state, the fingers domain adopts a more closed, or "clenched" conformation as compared to the Substrate Insertion Complex (Structure II), which is achieved by an additional rotation of the fingers domain by 13^0 and translation of the O/Y helices by 4 Å towards the template strand of DNA (Fig. 2e, f). This movement appears to push the template strand of DNA away from the active site, with the sharp 90^0 bending of the DNA induced between bases $n-1$ and n, unlike the bend between bases n and $n + 1$ in Structure II (Fig. 2b, e, f). This additional movement of the fingers domain results in rotating the n base out of the active site and brings the Y999 residue into the position normally occupied by substrate NTP in the Insertion Site, as evidenced by superimposing Structures IV and II using conserved residues in the palm subdomain

(Supplementary Movie 4). It is tempting to speculate that non-cognate NTP bound in the Entry Site and subsequently moved into the Insertion Site during fingers closing would be rejected from the EC upon fingers "clenching" in Structure IV. This is consistent with the observation that the high-quality cryo-EM map representing the Rejection Complex (Structure IV) was obtained only for the EC with a non-cognate substrate (Supplementary Fig. 2–8). Alternatively, Structure IV can represent a different post-translocated state of the EC, which has been observed previously in conditions when substrate NTP has not been provided[10].

Substrate transition from the entry to the Insertion Site and rejection of a noncognate substrate

Structural studies of the Pol A family of RNAPs, such as bacteriophage T7 and N4 RNAPs[1–3], and DNA polymerases, such as T7 DNAP, DNAP I, TAQ, and Pol Gamma[7,19,24,25], suggested that movement of the fingers domain from the "open" to "closed" conformation is responsible for delivery of the nucleotide into the catalytic site. Our data reveal the two most populated states of the mtRNAP EC with bound cognate substrate and demonstrate that the movement of the fingers domain delivers substrate NTP from the Entry site to the Insertion Site (Fig. 2c, Supplementary Movie 1-3). A two-pronged check-point mechanism of substrate selection by single-subunit RNA polymerases has been envisioned (Fig. 4)[1,2]. The initial stage of substrate selection involves formation of the insipient base-pair with the acceptor DNA base, observed in the crystal structure of T7 RNAP EC[2] and likely represented by multiple intermediate complexes on a pathway from the Entry to Insertion Site in mtRNAP in Complex ONE (Supplementary Fig. 2). The cognate substrate retained in the Pre-insertion Site travels into the Insertion Site where it is poised for a catalysis. When the non-cognate NTP is bound, imperfect base-paring results in incomplete closing of the fingers domain and substrate dissociation from RNAP. This state of the fingers, termed the "ajar" conformation, has been observed in structural studies of DNA polymerase I bound to a non-cognate substrate[26]. Because the ajar conformation remains relatively open, complete closure of the fingers domain is required to engage an
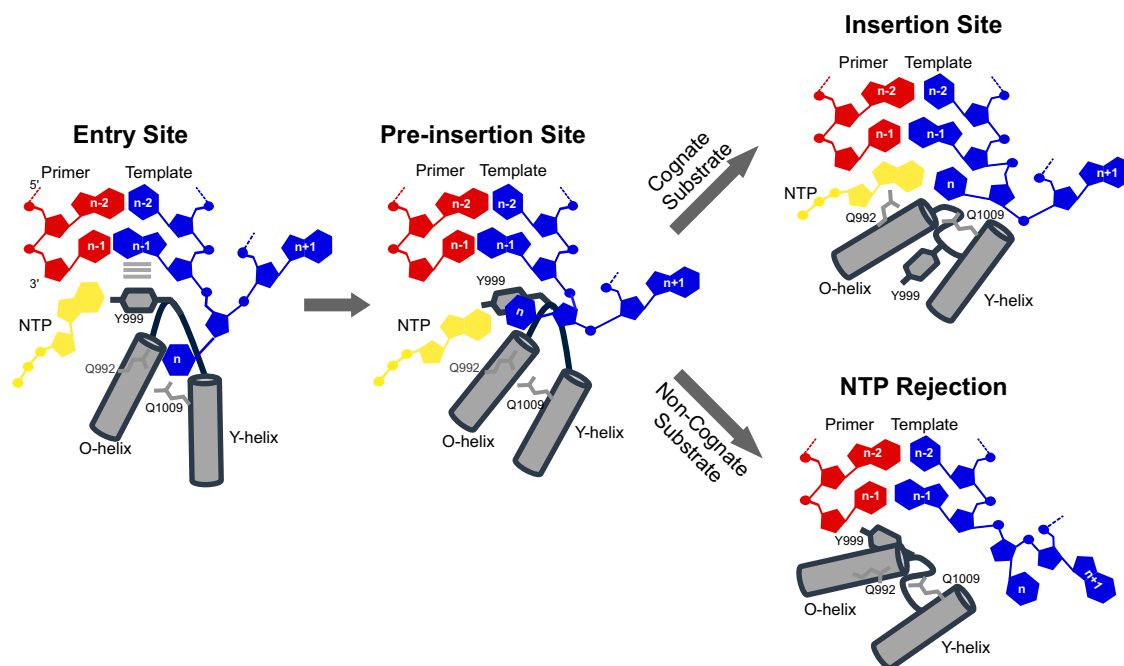
Fig. 4 | The mechanism of substrate binding and selection in Pol A family of RNA and DNA polymerases. Schematic illustration of the major states along nucleotide addition pathway. Left - substrate binding in the Entry Site observed in RNAPs and DNAPs (8U8U, 9BDD, 4YFU); center: substrate Pre-Insertion Site observed in RNAP (1S0V); top right - substrate Insertion Site (RNAPs and DNAPs: 8U8V; 1S76, 2HVI). Bottom right - substrate rejection complex (9BDC).

induced fit mechanism of substrate selection at the Insertion Site[27,28]. This mechanism relies on multiple local rearrangements of the active site residues, which assess the geometry of the base-pair[29,30] and the conformation of the sugar moiety, thereby enabling discrimination against dNTPs[31]. These rearrangements likely trigger additional movement of the fingers domain to expel the non-cognate nucleotide, facilitated by the Y999 residue of the O helix, ultimately resulting in the formation of the Substrate Rejection Complex observed within Complex TWO dataset (Fig. 2e, f). The model (Fig. 4) is based on experimental data and offers a unified view of substrate selection in single-subunit RNAPs and, possibly, in the Pol A family of DNAPs[24], explaining the previously observed substrate-bound intermediate states[1,2].

Translocation of motor enzymes is thought to involve a power stroke mechanism and/or a Brownian-ratchet mechanism[32]. The latter mechanism has been proposed for multi-subunit RNAPs[33,34], while the former – for single-subunit RNAPs. The structure of the T7 RNAP-product complex with the bound pyrophosphate[1] reveals that the formation of the phosphodiester bond does not trigger the enzyme's translocation along DNA. Rather, the release of the pyrophosphate results in the opening of the finger's domain and transfer of the 3′ end of RNA into the product site[1]. The lack of structural changes and heteroduplex translocation in mtRNAP EC-TEFM upon NTP binding in the Entry Site also argues in support of the power stroke mechanism in single-subunit RNAPs. On the other hand, sequestering of the *n* DNA base between O and Y helices prior to the initial NTP binding and moving Y999 residue in and out of the active site appear to represent the Brownian-ratchet mechanism of translocation, which ensures the unidirectional motion of polymerase during the NAC. Recent findings suggest that both the power stroke and Brownian-ratchet mechanisms are not mutually exclusive modes of motor enzyme translocation[32].

Methods
Protein expression and purification
N-terminal His-tagged variant of human mtRNAP (Δ119) and TEFM (Δ50) were expressed in *E.coli* BL21 (DE3) RIL cells (Agilent) and purified exactly as described previously[11]. Briefly, cells were grown until the OD600 reached 0.6 units, and expression was induced by addition of IPTG (0.15 mM) and carried out at 16°C for 18 h. MtRNAP and TEFM were purified by affinity chromatography using Ni-NTA beads (QIAGEN) followed by Heparin affinity chromatography using a HiTrap Heparin HP column (GE Healthcare). The proteins were further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated with 40 mM Tris/HCl pH 8.0, 300 mM NaCl, 20 mM MgCl₂, 5% Glycerol, 10 mM DTT). Peak fractions were collected and analyzed using PAGE electrophoresis. MtRNAP variants (Y999F, H1125A, Q992A, T996A, Q1009A and Y999F/H1125A) were obtained using site-directed mutagenesis (QuikChange, Agilent).

DNA and RNA oligonucleotides, and scaffold preparation
The sequences of the oligos (IDTDNA) are as follows (5′ to 3′): GGACATGGTGTAATTATTTCGACGCCA

GACGACC, NT27mt, GGTCGTCTGGCGTGCGCGCCGTTACACCA TGTCC, TS31mt, GGACATGGTGTAATTATTTCGACGCCAGACGACC, NT27mt + 1 T, AGUCUGCGGCGCGC, RNA14mt, and GGTCGTCTGGC GTGCGCGCCGTTACACCATGTCC, TS31mt + 1 A. Oligonucleotides were dissolved in water and mixed at a final concentration of 0.5 mM. The scaffolds were annealed by heating the mixture to 95°C for 7 min and cooled down stepwise (1°C /min) for 1 h to 25°C in a thermocycler.

Transcription assays
The activity of mtRNAP variants was assayed using the primer extension assay described previously[35]. RNA primers were labeled using

[γ-³²P] ATP (3000 Ci/mmol) and T4 Polynucleotide Kinase (NEB). The ability of mtRNAP variants to discriminate between ribo- and deoxy-NTPs was probed using a competition assay. ECs (0.5 μM) assembled using the NT27mt/TS31mt/R14 scaffold were incubated in the presence of rNTP/dNTP mixture containing 2 μM rATP/20 μM dATP (1:10 ratio), 2 μM rATP/200 μM dATP (1:100 ratio), and 2 μM rATP/2000 μM dATP (1:1,000 ratio) for 5 min at 37°C. Control reactions contained either rATP (100 μM) or dATP (100 μM). The reactions were carried out for 10 min at room temperature and stopped by the addition of an equal volume of a 2× stop solution (95% formamide/0.05 M EDTA). The products of the reaction were resolved by 20% PAGE containing 6 M urea and visualized by autoradiography using PhosphorImager (GE Healthcare).

Preparation of substrate-bound EC-TEFM complex for cryo-EM
Substrate-bound EC/TEFM complexes (5 μM) were assembled by incubating Δ119 mtRNAP with the NT27mt/TS31mt/R14 scaffold (Complex ONE) or NT27mt + 1 T/TS31mt + 1 A/R14 scaffold (Complex TWO) in buffer containing 20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM DTT, and 10 mM MgCl₂ for 10 min at room temperature followed by overnight dialysis at 4°C in the same buffer. After dialysis, α,β-methylene ATP (Sigma) was added to a final concentration of 1 mM. UltrAuFoil® R1.2/1.3, Gold 300 mesh grids (Quantifoil) were negatively glow-discharged with 25 mA for 60 s using a PELCO easiGlow™ Glow Discharge Cleaning System prior to preparation. Sample (3 μl) was then applied to grids that were blotted with PELCO® qualitative cellulose filter paper (Grade 595 55/20 mm) and vitrified with a Vitrobot Mark IV (ThermoFisher Scientific) for 5 s at 4 °C and 95-100 % humidity. The sample quality and particle distribution were assessed using EPU (Thermo Fisher Scientific) operating a 200 kV Glacios transmission electron microscope (ThermoFisher Scientific) equipped with a Falcon 4 direct electron detector.

Single-particle data acquisition and image processing
Data for Complex ONE (cognate substrate) and Complex TWO (non-cognate substrate) were collected at the National Cryo-Electron Microscopy Facility (NCEF) at NCI using a 300 kV Titan Krios transmission electron microscope (ThermoFisher Scientific) equipped with a Gatan K3 direct electron detector and a Gatan BioQuantum Energy Filter with a 20 eV slit width. Movies for Complex ONE were recorded in super-resolution mode using Latitude S (Gatan, Inc.) with a magnification of 105,000, corresponding to a pixel size of 0.428 Å over a defocus range of −0.5 to −2.25 μm. A dose rate of 12.93 e⁻/s/Physical Pixel resulted in a total electron dose of 60 e⁻/Å², which was fractionated into 40 frames. Movies for Complex TWO were recorded in Counting mode using Latitude S (Gatan, Inc.) with a magnification of 105,000, corresponding to a pixel size of 0.835 Å over a defocus range of −0.75 to −2.00 μm. A dose rate of 18.7 e⁻/s/Physical Pixel resulted in a total electron dose of 50.6 e⁻/Å², which was fractionated into 40 frames.

Workflows for image processing of substrate-bound EC-TEFM structures are found in Supplementary Fig. 2,5. The dose-fractionated movie stacks were imported and processed in cryoSPARC[36]. The movies were motion-corrected, gain-normalized, drift-corrected, summed, and dose-weighted using MotionCor2[37]. The super-resolution movies were Fourier cropped (to 0.856 Å). Contrast transfer function (CTF) values were estimated using patch CTF (cryoSPARC) or GCTF[38]. Micrographs with ice, ethane contamination, and/or poor CTF fit resolution were discarded. Initial particles were picked and extracted from the dose-weighted images with a box size of 296 px using the cryoSPARC blob picker with dimensions of 80-140 Å. The particles were aligned and sorted using several rounds of cryoSPARC 2D classification (n = 100) to generate an initial model using cryoSPARC ab initio reconstruction (n = 1),

resulting in a consensus model used for downstream refinement. Particles representing the EC-TEFM complex were further curated using the initial model as a 3D template for cryoSPARC non-uniform 3D refinement[39] and cryoSPARC heterogeneous refinement (n = 3), resulting in two classes used for further analysis and one 'junk' class that was excluded from further analysis.

To assess variability within each dataset, cryoSPARC 3D Variability Analysis (3DVA)[24] was performed (n = 5 modes, filter resolution limit 4 Å) followed by 3D Variability Display[40] (Intermediate output mode, n = 6 clusters).

For Complex ONE, Cluster 1 contained 368,317 particles and was curated by removing duplicate particles before performing a final non-uniform refinement to obtain Structure I with a nominal resolution of 2.90 Å. Cluster 6 contained 331,858 particles. Particles within Cluster 6 were further processed through Bayesian particle polishing[41] in RELION 4.0[42]. Polished particles were then imported back into cryoSPARC for a final non-uniform refinement, resulting in Structure II with a nominal resolution of 2.74 Å. Clusters 2-5 underwent a similar processing procedure, but were excluded from further analysis due to partial ATP density.

For Complex TWO, Clusters 1 and 2 resembled substrate-bound EC-TEFM and contained 497,782 particles. Particles were curated by removing duplicates before performing a final non-uniform refinement to yield Structure III with a nominal resolution of 2.86 Å. Clusters 3-6 resembled the "closed" EC-TEFM complex with no bound substrate and contained 2,457,845 particles. Particles were curated by removing duplicates before performing a final non-uniform refinement to yield Structure IV with a nominal resolution of 2.54 Å.

The reported resolutions of the cryo-EM maps are based on the Fourier shell correlation (FSC) 0.143 criterion[43]. Local resolution calculations were generated using cryoSPARC Local Resolution Estimation and displayed in ChimeraX[44] (Supplementary Fig. 3,4,6,7). The angular distribution of particle orientations and directional resolution through the 3DFSC package[45] (Supplementary Fig. 3,4,6,7) were calculated for all structures.

Model building and structure refinement

An initial model of the TEFM-mtRNAP complex was derived from Protein Data Bank (PDB) 5OLA[11]. The models were manually docked into the respective cryo-EM maps for each of the structures using USCF Chimera[44]. DNA-B and RNA-A restraints from Coot were used to fit the polynucleotide chains in the non-template DNA, template DNA, and RNA primer strands. The non-hydrolyzable ATP substrate, α,β-methylene ATP, was fit using Jiggle-Fit Ligand in Coot v.0.9.8.2[46]. The models were Rigid-body and Real-Space Refined using PHENIX real-space-refine[47]. For the real-space refinement, rigid-body refinement was carried out with Ramachandran and secondary structure restraints. Models were inspected and modified in Coot, and the refinement process was repeated iteratively to obtain the final structures. Comprehensive model validation was carried out with PHENIX[48] and the PDB validation server (https://validate-rcsb-2.wwpdb.org/) and is summarized in Supplementary Tables 1,2. Figures and movies were generated with PYMOL and ChimeraX.

Generation of the sequence logos for O/Y helices of mitochondrial and phage RNAPs

DNA sequence logos were built using the weblogo.berkeley.edu/logo.cgi server. MtRNAPs from mammalian, avian, and bony fish species (*H. sapiens, Pan troglodytes, Pan paniscus, Nomascus leucogenys, Papio Anubis, Macaca mulata, Callithrix jacchus, Saimiri boliviensis, Otolemur garnettii, A. melanoleuca, Canis lupus familiaris, Sus scrofa, Loxodonta Africana, Camelus dromedarius, Camelus bactrianus, Lipotex vexillifer, Mustela putorius, Bos taurus, Capra hirscus,*

Mus musculus, Cricetulus griseus, Rattus norvegicus, Heterocephalus glaber, Cavia porcellus, Monodelphis domestica) were used to generate the sequence logo for O/Y helices of mtRNAP. Sequences of RNAPs from T7- and N4-like phages (VP93, PT2, PT5, LKA1, LKD16, phiKMV, phikF77, Era103, K1E, SP6, K1-5, phiSG-JL2, Berlin, VP4, phiA1122, gh-1, phiYeO3-12, Kvp1, T7, T3, Yepe2, K11, 13a, EcoDS1, BA14, MmP1, Syn5, P-SSP7, P60, N4, KP32, Xop411, OP1, and Xp10) were used for O/Y helices logo.

Statistics and reproducibility

Experiments presented in Fig. 3b, c, e and Supplementary Figs. 9a, b, c were repeated at least three times. The representative gel images are shown.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM maps and atomic coordinates were deposited in the Electron Microscopy Data Bank (https://www.ebi.ac.uk/emdb) under accession codes EMD-42027, EMD-42028, EMD-44449, and EMD-44448 and in the Protein Data Bank under accession codes 8U8U, 8U8V, 9BDD, and 9BDC. Previously published protein structure data used for analysis in this study are available in the Protein Data Bank (www.rcsb.org) under PDB ID: 4BOC, 5OLA (human mitochondrial RNAP EC), 1S0V, 1S76 (T7 RNAP EC), 2HVI, 4YFU *Geobacillus stearothermophilus* DNAP. Source data are provided with this paper.

References

1. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
2. Temiakov, D. et al. Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).
3. Gleghorn, M. L., Davydova, E. K., Basu, R., Rothman-Denes, L. B. & Murakami, K. S. X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc. Natl Acad. Sci. USA* **108**, 3566–3571 (2011).
4. Basu, R. S. & Murakami, K. S. Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J. Biol. Chem.* **288**, 3305–3311 (2013).
5. Jackson, N. A. C., Kester, K. E., Casimiro, D., Gurunathan, S. & DeRosa, F. The promise of mRNA vaccines: a biotech and industrial perspective. *NPJ Vaccin* **5**, 11 (2020).
6. Bonekamp, N. A. et al. Small-molecule inhibitors of human mitochondrial DNA transcription. *Nature* **588**, 712–716 (2020).
7. Park, J., Baruch-Torres, N. & Yin, Y. W. Structural and Molecular Basis for Mitochondrial DNA Replication and Transcription in Health and Antiviral Drug Toxicity. *Molecules* **28**, 1796 (2023).
8. Goovaerts, Q. et al. Structures illustrate step-by-step mitochondrial transcription initiation. *Nature*, (2023).
9. Ringel, R. et al. Structure of human mitochondrial RNA polymerase. *Nature* **478**, 269–273 (2011).
10. Schwinghammer, K. et al. Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* **20**, 1298–1303 (2013).
11. Hillen, H. S. et al. Mechanism of Transcription Anti-termination in Human Mitochondria. *Cell* **171**, 1082–1093.e13 (2017).
12. Hillen, H. S., Morozov, Y. I., Sarfallah, A., Temiakov, D. & Cramer, P. Structural Basis of Mitochondrial Transcription Initiation. *Cell* **171**, 1072–1081.e10 (2017).
13. Johnson, S. J., Taylor, J. S., Beese, L. S. & Processive, D. N. A. synthesis observed in a polymerase crystal suggests a mechanism

for the prevention of frameshift mutations. *Proc. Natl Acad. Sci. USA* **100**, 3895–3900 (2003).

14. Oh, J. et al. Structural basis of transcription recognition of a hydrophobic unnatural base pair by T7 RNA polymerase. *Nat. Commun.* **14**, 195 (2023).

15. Miller, B. R. 3rd, Beese, L. S., Parish, C. A. & Wu, E. Y. The Closing Mechanism of DNA Polymerase I at Atomic Resolution. *Structure* **23**, 1609–1620 (2015).

16. Warren, J. J., Forsberg, L. J. & Beese, L. S. The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. *Proc. Natl Acad. Sci. USA* **103**, 19701–19706 (2006).

17. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, 251–258 (1998).

18. Szymanski, M. R. et al. Structural basis for processivity and antiviral drug toxicity in human mitochondrial DNA replicase. *EMBO J.* **34**, 1959–1970 (2015).

19. Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525 (1998).

20. Tunitskaya, V. L. & Kochetkov, S. N. Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochem. (Mosc.)* **67**, 1124–1135 (2002).

21. Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269**, 25120–25128 (1994).

22. Kostyuk, D. A. et al. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Lett.* **369**, 165–168 (1995).

23. Huang, Y., Eckstein, F., Padilla, R. & Sousa, R. Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* **36**, 8231–8242 (1997).

24. Rothwell, P. J. & Waksman, G. Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **71**, 401–440 (2005).

25. Beese, L. S., Friedman, J. M. & Steitz, T. A. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32**, 14095–14101 (1993).

26. Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286**, 19758–19767 (2011).

27. Joyce, C. M. & Benkovic, S. J. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43**, 14317–14324 (2004).

28. Wong, I., Patel, S. S. & Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526–537 (1991).

29. Jamsen, J. A., Shock, D. D. & Wilson, S. H. Watching right and wrong nucleotide insertion captures hidden polymerase fidelity checkpoints. *Nat. Commun.* **13**, 3193 (2022).

30. Radhakrishnan, R. et al. Regulation of DNA repair fidelity by molecular checkpoints: "gates" in DNA polymerase beta's substrate selection. *Biochemistry* **45**, 15142–15156 (2006).

31. Makinen, J. J. et al. The mechanism of the nucleo-sugar selection by multi-subunit RNA polymerases. *Nat. Commun.* **12**, 796 (2021).

32. Hwang, W. & Karplus, M. Structural basis for power stroke vs. Brownian ratchet mechanisms of motor proteins. *Proc. Natl Acad. Sci. USA* **116**, 19777–19785 (2019).

33. Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).

34. Belogurov, G. A. & Artsimovitch, I. The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J. Mol. Biol.* **431**, 3975–4006 (2019).

35. Sarfallah, A. & Temiakov, D. In Vitro Reconstitution of Human Mitochondrial Transcription. *Methods Mol. Biol.* **2192**, 35–41 (2021).

36. 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).

37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

38. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

39. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

40. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).

41. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

42. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J.* **478**, 4169–4185 (2021).

43. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).

44. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

45. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

46. Casanal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).

47. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D. Struct. Biol.* **74**, 531–544 (2018).

48. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D. Biol. Crystallogr* **66**, 213–221 (2010).

Acknowledgements

Author contributions

Experimental design and conceptualization: K.H., A.R.N., D.T. Protein preparation and biochemical experiments: K.H. Cryo-EM Data acquisition: K.H. Cryo-EM Single particle analysis and model building: K.H., A.R.N. Writing – original draft: D.T. Writing – review & editing: D.T., K.H., A.R.N.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-50817-9.

Correspondence and requests for materials should be addressed to Dmitry Temiakov.

Peer review information *Nature Communications* thanks Dong Wang, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.