Doctoral Thesis

Piotr Purzycki

Structural basis for functional cooperation between proteins involved in the processing of RNA primers during mitochondrial DNA replication

Supervisor: dr hab. Michał R. Szymański, prof. UG

Laboratory of Structural Biology
Intercollegiate Faculty of Biotechnology UG&MUG

Gdańsk 2025

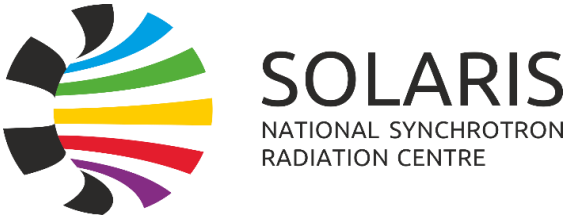
This doctoral dissertation was made possible thanks to funding from the following projects:

1. The EMBO Installation Grant (No. 4129) awarded to prof. Michał R. Szymański by the European Molecular Biology Organization

2. The Polish National Science Centre Grant (2022/46/E/NZ1/00430) awarded to prof. Michał R. Szymański.

3. I gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2025/018735

4. I would like to acknowledge the National Synchrotron Radiation Centre SOLARIS in Kraków, Poland, for providing access to its research infrastructure. Research at the National Synchrotron Radiation Centre SOLARIS is supported by the Ministry of Science and Higher Education, Poland, under contract no. 1/SOL/2021/2.



I would like to express my gratitude to Dr hab. Michał R. Szymański, prof. UG, for his supervision, guidance, and for providing the opportunity to carry out this project. His support was essential for the completion of this thesis.

I further acknowledge the present and former members of the Structural Biology Laboratory as well as the academic, technical, and administrative personnel of the faculty, for their assistance and for ensuring a well-organised and safe environment in which this research was carried out.

Finally, I would like to acknowledge all others who, in any capacity, contributed to the completion of this project.

Table of content

1. Streszczenie

Ludzkie mitochondrialne DNA (mtDNA) to kolista, dwuniciowa cząsteczka DNA o długości 16,6 kilopar zasad, której replikacja i utrzymanie są niezbędne dla funkcjonowania komórki; mutacje w mtDNA prowadzą do ciężkich chorób. System replikacji mtDNA różni się od replikacji jądrowej, poprzez wykorzystanie białek dedykowanych mitochondriom. W powszechnie cytowanym modelu wypierania nici, synteza nici wiodącej rozpoczyna się na początku replikacji nici ciężkiej. Gdy dwie trzecie genomu zostanie zreplikowane, odsłaniany jest początek replikacji nici lekkiej, tworząc strukturę pętli typu stem-loop. Po zakończeniu replikacji startery RNA na końcach 5′ nowo zsyntetyzowanych nici muszą zostać usunięte. Defekty w tym procesie mogą prowadzić do powstawania patogennych cząsteczek mtDNA, skutkując różnorodnymi chorobami mitochondrialnymi. Ostatnie badania dotyczące terminacji replikacji mtDNA zidentyfikowały funkcjonalne interakcje między rybonukleazą H1 (RNaza H1), mitochondrialną egzonukleazą 5′ G (EXOG) i polimerazą DNA Pol γ, jednak dokładny mechanizm usuwania starterów RNA w ludzkich mitochondriach pozostaje jedynie częściowo zdefiniowany. W celu zbadania molekularnych podstaw dwóch odrębnych interakcji z RNazą H1, niniejsza rozprawa ma na celu wyjaśnienie funkcjonalnego oddziaływania RNazy H1 z EXOG w procesowaniu końcowym primerów RNA oraz współdziałania RNazy H1 z Pol γ podczas replikacji. W badaniach wykorzystano testy biochemiczne i kinetyczne, w tym powierzchniowy rezonans plazmonowy (SPR) i interferometrię biowarstwową (BLI), analizę mutacyjną, modelowanie strukturalne za pomocą AlphaFold, krystalografię rentgenowską oraz mikroskopię krioelektronową (cryo-EM).

W pierwszej części projektu, zbadałem funkcjonalne oddziaływanie między RNazą H1 a EXOG. Za pomocą SPR zmapowałem główną powierzchnię interakcji w domenie katalitycznej RNazy H1. W biochemicznej rekonstytucji potwierdziłem, że ta interakcja jest istotna funkcjonalnie, gdyż do całkowitego wycięcia starterów RNA *in vitro* wymagane były oba, wspólnie działające enzymy, czego nie udało się osiągnąć przez żadne z białek osobno. Ponadto EXOG była w stanie częściowo przywrócić upośledzoną aktywność chorobotwórczego wariantu RNazy H1 co może sugerować znaczenie kliniczne. W celu zbadania strukturalnych podstaw tego oddziaływania zastosowałem zintegrowane podejście, łączące próby krystalizacji z predykcjami AlphaFold2 Multimer. Chociaż krystalografia nie pozwoliła na rozwiązanie struktury kompleksu, modelowanie komputerowe zidentyfikowało kluczowe reszty na powierzchni interakcji, które następnie zweryfikowałem za pomocą mutagenezy ukierunkowanej i testów aktywności biochemicznej.

W drugiej części rozprawy badałem współdziałanie RNazy H1 z Pol γ. W eksperymentach SPR pokazałem bezpośrednią interakcję pomiędzy obydwoma białkami, w której pośredniczy domena katalityczna RNazy H1. Uzyskane przeze mnie wyniki pokazują, że synteza DNA poprzez Pol γ stymuluje RNaze H1 do całkowitego usunięcia starterów. Ponadto, katalitycznie nieaktywny wariant RNazy H1 zatrzymywał Pol γ w trakcie syntezy DNA, działając jako fizyczna przeszkoda. Podczas gdy dowody biochemiczne wspierają istnienie trójskładnikowego kompleksu, mikroskopia cryo-EM wykazała heterogenność konformacyjną, która uniemożliwiła wysokorozdzielczą wizualizację RNazy H1. Aby rozwiązać ten problem, wykonałem predykcję za pomocą AlphaFold3, co umożliwiło mi uzyskanie wiarygodnego modelu kompleksu zgodnego z danymi biochemicznymi. Dane strukturalne i biochemiczne zasugerowały określoną organizację przestrzenną, która umożliwia koordynację podczas terminacji replikacji mtDNA. Łącznie, wyniki te dostarczyły dowodów na istnienie komplementarnych mechanizmów, za pomocą których RNaza H1 koordynuje działanie z różnymi partnerami enzymatycznymi, aby zapewnić całkowite usunięcie starterów RNA i terminację replikacji w ludzkich mitochondriach.

Niniejsza praca określa specyficzne warunki biochemiczne, w których RNaza H1 jest zdolna do całkowitego cięcia startera RNA *in vitro*, dając podstawy do głębszego zrozumienia dotychczas niewyjaśnionych procesów. Integrując dowody strukturalne, biochemiczne i kinetyczne, niniejsza rozprawa poszerza dostępną wiedzę w zakresie dojrzewania mitochondrialnych starterów RNA i podkreśla, w jaki sposób aktywność RNazy H1 przyczynia się do terminacji replikacji i stabilności genomu mitochondrialnego.

2. Abstract

Human mitochondrial DNA (mtDNA) is a circular, double-stranded molecule of 16.6 kilobase pairs whose replication and maintenance are essential for cellular function, as mutations in mtDNA cause severe mitochondrial diseases. mtDNA replication differs fundamentally from nuclear replication and relies on a dedicated set of mitochondrial proteins. In the widely cited strand displacement model, leading strand synthesis initiates at the heavy-strand origin. When approximately two-thirds of the genome have been replicated, the light-strand origin becomes exposed and adopts a stem-loop structure. After replication is completed, RNA primers at the 5′ ends of nascent strands must be removed. Defects in this step lead to the formation of pathogenic mtDNA species and contribute to diverse mitochondrial disorders. Although recent studies have identified functional interactions among human mitochondrial ribonuclease H1 (RNase H1), the mitochondrial 5′-exonuclease G (EXOG), and DNA polymerase gamma (Pol γ), the precise mechanism of RNA primer removal in human mitochondria remains only partially understood. To address this gap, I investigated two distinct RNase H1-centred interactions: its cooperation with EXOG in terminal RNA primer processing and its functional interplay with Pol γ during mtDNA replication. The work integrates biochemical and kinetic assays, including surface plasmon resonance (SPR) and biolayer interferometry (BLI), mutational analysis, structural modelling using AlphaFold, X ray crystallography, and cryo electron microscopy (cryo-EM).

In the first part of this project, I examined the functional cooperation between RNase H1 and EXOG. SPR analysis mapped the primary interaction surface to the catalytic domain of RNase H1. Biochemical reconstitution demonstrated that both enzymes are required together to achieve complete excision of RNA primers *in vitro*, a result that neither protein can achieve on its own. Furthermore, EXOG partially restored the impaired activity of a disease associated RNase H1 variant, suggesting potential clinical relevance. To investigate the structural basis of this cooperation, I applied an integrated approach combining crystallisation attempts with AlphaFold2 Multimer predictions. While crystallography did not yield a complex structure, computational modelling identified key interfacial residues that were subsequently validated through site directed mutagenesis and biochemical activity assays.

The second part of the thesis explored the interplay between RNase H1 and Pol γ. SPR experiments revealed a direct interaction mediated by the catalytic domain of RNase H1. This association enabled a reciprocal functional relationship. DNA synthesis by Pol γ stimulated RNase H1 to achieve complete primer removal, whereas a catalytically inactive RNase H1 variant halted Pol γ during gap-filling synthesis, acting as a physical obstacle. Although biochemical data supported the existence of a ternary assembly, cryo-EM analysis revealed

substantial conformational heterogeneity that prevented high resolution visualisation of RNase H1. To address this limitation, AlphaFold3 modelling generated a plausible structural model consistent with the biochemical observations and suggested a spatial arrangement that could facilitate coordinated primer removal during mtDNA replication termination.

Together, these findings reveal complementary mechanisms through which RNase H1 collaborates with distinct enzymatic partners to ensure complete RNA primer removal and proper replication termination in human mitochondria. This work also defines the specific biochemical conditions under which RNase H1 is capable of complete RNA primer cleavage *in vitro*, clarifying previously unresolved observations. By integrating structural, biochemical, and kinetic evidence, the thesis expands current understanding of mitochondrial RNA primer maturation and highlights the contribution of RNase H1 activity to replication termination and mitochondrial genome stability.

3. List of abbreviations

Å	Angstrom
A280	Absorbance at 280 nm
BLI	Biolayer interferometry
BSA	Bovine serum albumin
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cryo-EM	Cryo electron microscopy
CSB2	Conserved sequence block 2
CTD	C-terminal catalytic domain
D loop	Displacement loop
DNA	Deoxyribonucleic acid
DNA2	DNA replication helicase nuclease 2
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EndoG	Endonuclease G
EXOG	Exonuclease G
FEN1	Flap endonuclease 1
HBD	N-terminal hybrid-binding domain
HBS EP	HEPES buffered saline with EDTA and surfactant
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSP	Heavy-strand promoter
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ipTM	Inter-chain predicted TM score
K_d	Equilibrium dissociation constant
kDa	Kilodalton
k_{off}	Dissociation rate constant
k_{on}	Association rate constant
LB	Lysogeny broth
Lig III	DNA ligase III
LSP	Light-strand promoter
MGME1	Mitochondrial genome maintenance exonuclease 1
MST	Microscale thermophoresis
MTS	Mitochondrial targeting sequence
mtSSB	Mitochondrial single stranded DNA binding protein
MWCO	Molecular weight cut off
NCR	Non-coding region
NP-40	Nonidet P 40
NTA	Nitrilotriacetic acid
OD	Optical density
OD600	Optical density at 600 nm
OriH	Origin of heavy strand replication
OriL	Origin of light strand replication
PAE	Predicted aligned error
PAGE	Polyacrylamide gel electrophoresis

PDB	Protein databank
PLA	Proximity ligation assay
pLDDT	Predicted local distance difference test
PMSF	Phenylmethylsulfonyl fluoride
Pol γ	DNA polymerase gamma holoenzyme
Pol γA	DNA polymerase gamma catalytic subunit A
Pol γB	DNA polymerase gamma accessory subunit B
POLG	DNA polymerase subunit gamma 1 gene
POLG2	DNA polymerase subunit gamma 2 gene
POLRMT	Mitochondrial RNA polymerase
pTM	Predicted TM score
R loop	RNA/DNA hybrid three stranded structure
RH1	Ribonuclease H1
RITOLS	Ribonucleotide incorporation throughout the lagging strand
RNA	Ribonucleic acid
RNase H1	Ribonuclease H1
RNASEH1	Ribonuclease H1 gene
RPA	Replication protein A
RPM	Rotation per minute
RU	Response units
SCD	Strand coupled DNA synthesis model
SDM	Strand displacement model
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
Sf9	Spodoptera frugiperda 9 cells
SPR	Surface plasmon resonance
ssDNA	Single stranded DNA
TAS	Termination associated sequence
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEFM	Transcription elongation factor of mitochondria
TFAM	Transcription factor A mitochondrial
T_i	Inflection temperature
T_m	Melting temperature
TWINKLE	Mitochondrial DNA helicase Twinkle
TWNK	Twinkle helicase gene
UV	Ultraviolet
WT	Wild-type
ε_{280}	Extinction coefficient at 280 nm

4. Introduction

4.1. Mitochondria and mitochondrial DNA

Mitochondria are essential organelles of eukaryotic cells, primarily responsible for ATP production through oxidative phosphorylation, OXPHOS (Spinelli & Haigis, 2018). The majority of cellular ATP is generated by the OXPHOS system, which consists of five multi-subunit enzyme complexes located at the inner mitochondrial membrane. Beyond energy metabolism, mitochondria contribute to various cellular processes, including apoptosis, calcium homeostasis, and reactive oxygen species (ROS) signalling. Unique among organelles, mitochondria possess their own genome, which is inherited maternally and maintained separately from nuclear DNA (Gray, 2012).

The evolutionary origin of mitochondria is widely attributed to an ancient endosymbiotic event, in which a proto-eukaryotic host cell incorporated an ancestral alpha-proteobacterium (Gray, 2012; Martin et al., 2015). Although most genes of the original endosymbiont were either lost or transferred to the nucleus, the mitochondrion retained a compact genome that remains essential for organelle and cell viability (Gray, 2012). Human mitochondrial DNA (mtDNA) is a highly reduced, circular, double-stranded, multicopy molecule of 16.569 base pairs (Figure 1). It encodes 13 polypeptides of the respiratory chain, along with 22 transfer RNAs and 2 ribosomal RNAs (Calvo et al., 2016). The RNA components are required for intramitochondrial translation of the proteins encoded by mtDNA.

There are two strands of mtDNA, heavy (H) and light (L), that differ in base composition, making it possible to separate them by density gradient centrifugation (Berk & Clayton, 1974). The regulatory elements required to initiate replication and transcription are concentrated in a single intergenic non-coding region (NCR) (Figure 1). In humans, this region is about 1,100 base pairs long and contains three strand-specific promoters for transcription of mtDNA, namely two light-strand promoters LSP and LSP2 and one heavy-strand promoter HSP (Tan et al., 2022). In the nucleus DNA replication starts at defined origins and proceeds in two directions, generating leading and lagging strand synthesis (Kornberg & Baker, 1992). In contrast, the mitochondrial genome uses two strand-specific unidirectional origins of replication, one for synthesis of the leading H strand (OriH) and one for the lagging L strand (OriL). OriH is located within the NCR, whereas OriL lies in a cluster of tRNA genes approximately 11.000 base pairs downstream of OriH (Figure 1) (Robberson et al., 1972).

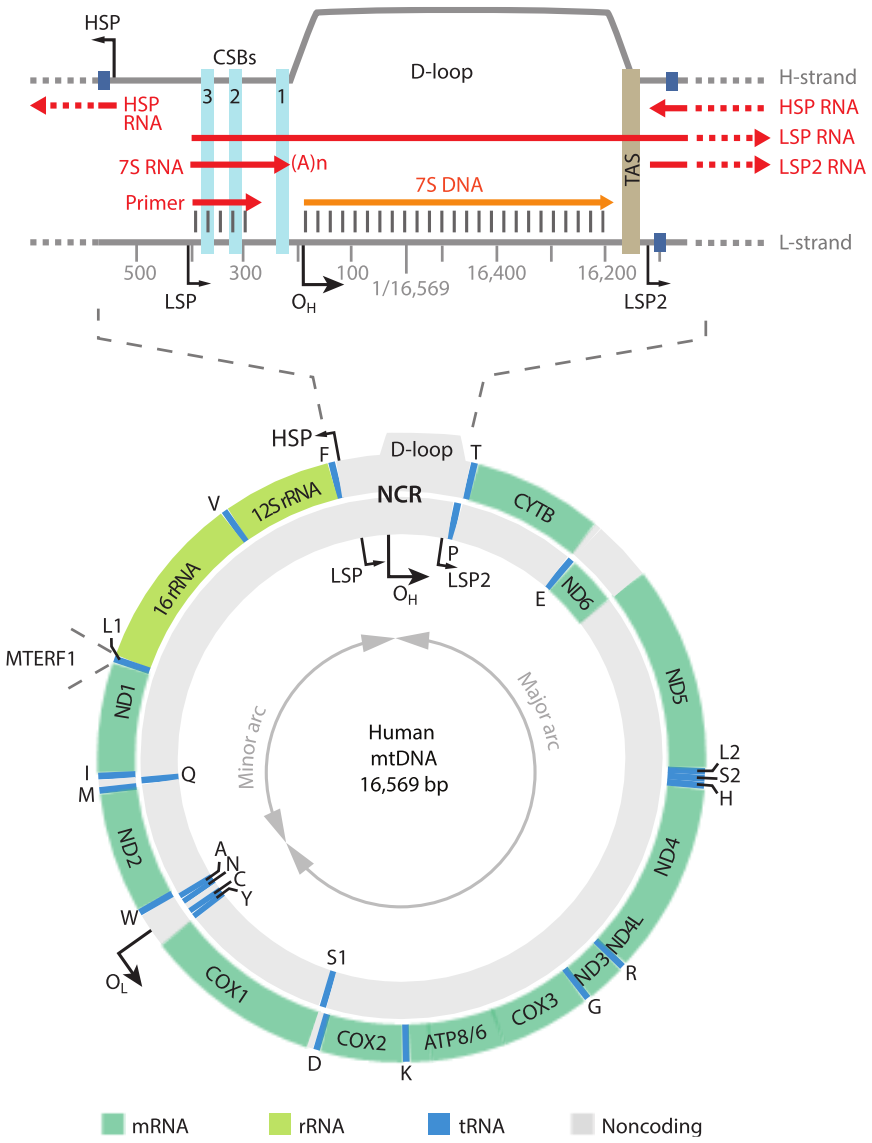


Figure 1. The human mitochondrial genome showing gene arrangement and regulatory regions. The circular, double-stranded DNA molecule of 16.569 base pairs is shown with the arrangement of its 37 genes, which encode 13 polypeptides (mRNAs), 22 transfer RNAs (tRNAs), and 2 ribosomal RNAs (rRNAs). The genome contains two strand-specific origins of replication, OriH and OriL, which divide it into a major and a minor arc. An enlarged view of the main non-coding region (NCR) is presented at the top, detailing key regulatory elements including the three promoters (HSP, LSP, and LSP2) and the origin of heavy-strand replication. This region also contains the characteristic triple-stranded displacement loop (D-loop), formed by the short replication product known as 7S DNA. (adapted from Falkenberg & Gustafsson, 2024).

Mitochondrial DNA is packaged into specialised nucleoprotein complexes referred to as nucleoids (Alam, 2003; Bogenhagen et al., 2008; Farge & Falkenberg, 2019; Satoh, 1991). The transcription factor TFAM is the principal structural component of nucleoids and is indispensable for the initiation of transcription (Falkenberg et al., 2024). By binding to mtDNA, TFAM bends and distorts the double helix, resulting in its compaction (Farge et al., 2012;

Kaufman et al., 2007; Kukat et al., 2011; Kukat et al., 2015). The degree of this packaging regulates the accessibility of promoter and origin regions, thereby influencing both replication and transcription (Kaufman et al., 2007; Farge et al., 2014). In addition to TFAM, nucleoids contain several other associated proteins, including replication machinery (DNA polymerase, helicase, and single-stranded DNA-binding protein), transcription factors (POLRMT, TFB2M), and regulatory elements such as LonP1, ATAD3, and mTERF family members, which collectively couple mtDNA maintenance with mitochondrial metabolism (Lee & Han, 2017).

4.2. Replication machinery

mtDNA replication depends on a different set of proteins from nuclear DNA replication (Shutt & Gray, 2006). The minimal replisome consists of three components. The central enzyme is DNA polymerase gamma (Pol γ). Pol γ is the sole DNA polymerase in mammalian mitochondria, responsible for both the replication and repair of the mitochondrial genome (mtDNA) (Gray & Wong, 1992; Hance et al., 2005; Kaguni, 2004). As a member of the high-fidelity DNA Polymerase A family, it shares ancestry with enzymes such as *E. coli* Pol I and bacteriophage T7 DNA polymerase (Braithwaite & Ito, 1993; Delarue et al., 1990). The functional holoenzyme is a heterotrimer, comprising a 140 kDa catalytic subunit, Pol γA, and a homodimer of 55 kDa accessory subunits, Pol γB, encoded by the *POLG* and *POLG2* genes, respectively (Wong et al., 2008; Chan & Copeland, 2009; Young et al., 2015; Lee et al., 2009).

The Pol γA subunit contains the core enzymatic functionalities within three distinct domains: a polymerase (pol) domain, a 3′-5′ proofreading exonuclease (exo) domain, and a large, intervening spacer domain unique to Pol γA (Figure 2) (Gray & Wong, 1992; Fan et al., 2006; Hance et al., 2005; Kaguni, 2004; Lee et al., 2009; Szymanski et al., 2015; Yakubovskaya et al., 2006). The pol domain adopts the canonical ‘right-hand’ fold (fingers, palm, and thumb) responsible for DNA synthesis. The palm subdomain houses the essential catalytic aspartate residues (D890 and D1135), while the fingers form the nucleotide-binding site crucial for substrate selection (Lee et al., 2009; Sohl et al., 2015; Szymanski et al., 2015). The exo domain provides a crucial proofreading function that ensures high fidelity, with an error frequency of less than $1x10^{-6}$ per incorporated nucleotide (Longley et al., 2001). The unique spacer domain, absent in other Pol I family members, is structurally divided into two functional subdomains: the Intrinsic Processivity (IP) subdomain, which enhances DNA binding on its own, and the Accessory Interacting Determinant (AID) subdomain, which mediates the critical interaction with Pol γB (Lee et al., 2009; Szymanski et al., 2015).

The Pol γB accessory subunit, while lacking intrinsic catalytic activity, is essential for high processivity (Johnson & Johnson, 2001; Lee et al., 2010). It functions as an allosteric regulator, enhancing Pol γA affinity for DNA and accelerating the rate of synthesis (Johnson & Johnson, 2001; Lee et al., 2010; Carrodeguas et al., 2002; Farge et al., 2007). Structural analyses reveal that the processivity of the holoenzyme is achieved through an asymmetric interaction, where Pol γA AID subdomain primarily engages one Pol γB monomer via extensive hydrophobic interactions. This binding induces a conformational change that exposes a lysine-rich region known as the 'K-tract' within the AID subdomain. This K-tract interacts electrostatically with the DNA backbone, extending the enzyme's contact footprint from approximately 10 base pairs to around 25 base pairs, thereby anchoring the holoenzyme to the template for prolonged synthesis (Lee et al., 2009).

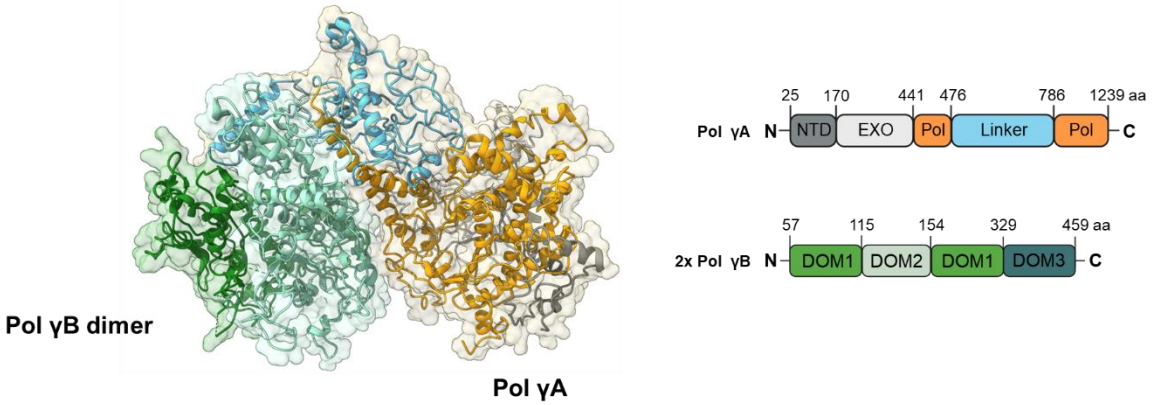


Figure 2. Domain organisation of Pol γA, and Pol γB. Apo structure of the Pol γ holoenzyme, with Pol γA in orange and Pol γB in green (PDB: 3IKM (Lee et al., 2009)). Structure was rendered in UCSF ChimeraX (Pettersen et al., 2021). Below: Schematic representation of the Pol γA and Pol γB domain structures with marked domain boundaries.

Given this intricate structural and functional architecture, it is unsurprising that over 300 pathogenic mutations in *POLG* and *POLG2* are a major cause of human mitochondrial diseases, as they disrupt catalysis, DNA binding, or the critical subunit interface, leading to replication stalling and mtDNA instability (Wong et al., 2008; Chan & Copeland, 2009; Young et al., 2015; Rahman & Copeland, 2019).

Pol γ is unable to replicate double-stranded DNA on its own and requires the helicase activity of TWINKLE (Spelbrink et al., 2001; Korhonen et al., 2004). This hexameric helicase unwinds the DNA double helix at the replication fork, providing a single-stranded template for Pol γ (Korhonen et al., 2004; Korhonen et al., 2003; Korhonen et al., 2008). The exposed single-

stranded DNA is immediately protected by the mitochondrial single-stranded DNA-binding protein (mtSSB), which prevents nuclease degradation and the formation of secondary structures (Mignotte et al., 1985; Miralles Fusté et al., 2014; Tiranti et al., 1993). These three proteins, Pol γ, TWINKLE, and mtSSB, form a highly coordinated complex where mtSSB stimulates the activities of both TWINKLE and Pol γ to ensure efficient and processive DNA synthesis (Korhonen et al., 2004; Oliveira & Kaguni, 2011).

A defining feature of this system is the mechanism of initiation. Unlike the nucleus, mammalian mitochondria lack a dedicated DNA primase (Chang et al., 1985a; Chang & Clayton, 1985b; Wanrooij et al., 2008; Fusté et al., 2010; Kuhl et al., 2016). This essential function is instead performed by the mitochondrial RNA polymerase (POLRMT). POLRMT synthesises short RNA transcripts that serve as primers to initiate DNA synthesis by Pol γ at both the OriH and OriL origins (Chang et al., 1985a; Chang & Clayton, 1985b; Fusté et al., 2010; Kuhl et al., 2016; Tsurumi & Lehman, 1990; Wanrooij et al., 2012; Wanrooij et al., 2008). This direct involvement of the transcriptional machinery in the replication process is a hallmark of the mitochondrial system, creating an intrinsic link between mtDNA expression and its maintenance. The necessity of RNA primers for initiation is a fundamental aspect common to all proposed models of mtDNA replication (Figure 3).

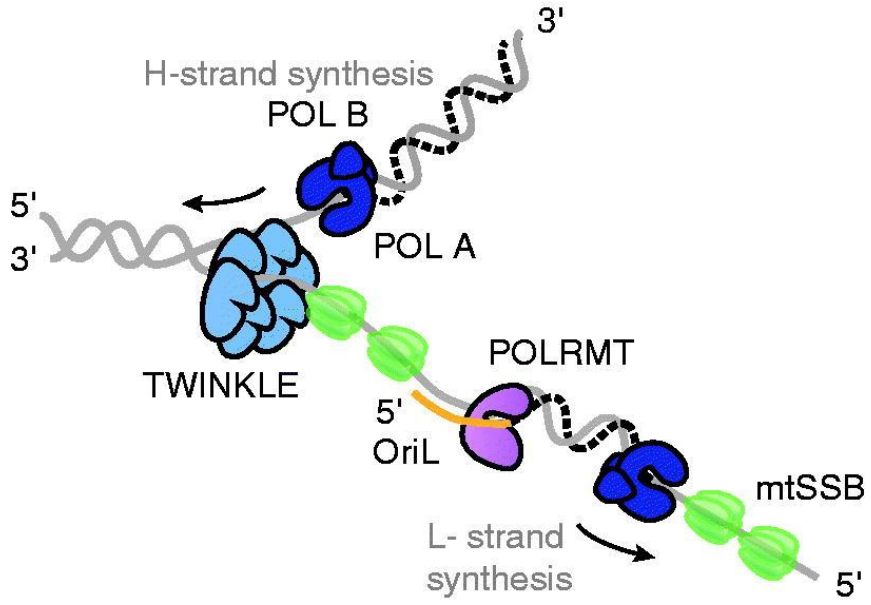


Figure 3. A schematic of the mitochondrial replication fork. The TWINKLE helicase (light blue) translocates on the parental H-strand in the 5′→3′ direction, unwinding the duplex ahead of the polymerase complex. mtSSB (green) binds and stabilises exposed single-stranded DNA and potentiates Pol γ-dependent synthesis of the nascent H-strand. Pol γ (dark blue; catalytic Pol γA plus accessory Pol γB) also performs L-strand synthesis using the displaced parental H-strand as template. POLRMT (purple) produces the short RNA primer (orange) required to initiate L-strand synthesis at OriL. (adapted from Falkenberg & Gustafsson, 2020).

4.3. Replication models

The precise manner in which the mitochondrial replisome proceeds to replicate the circular genome has been the subject of considerable debate, leading to the development of several distinct models supported by different lines of experimental evidence.

The longest-standing and most widely cited is the strand displacement model (SDM) (Figure 4) (Clayton, 1991; Robberson et al., 1972). Based on early electron microscopy studies, this model proposes an asynchronous and unidirectional mode of replication (Anderson et al., 1981). Synthesis begins at OriH, where a new H-strand is synthesised, displacing the parental H-strand in a structure known as a displacement loop (D-loop). This displaced strand remains single-stranded and coated by mtSSB for a significant portion of the replication cycle to prevent the POLRMT from initiating RNA synthesis on the displaced strand. Only when the replication fork has passed approximately two-thirds of the genome does it expose OriL. At this point, the parental H-strand folds into a stem-loop structure that prevents mtSSB binding and leaves the loop region accessible for POLRMT to initiate RNA synthesis (Miralles Fusté et al., 2014; Wanrooij et al., 2008; Fusté et al., 2010). This event triggers the initiation of L-strand synthesis in the opposite direction (Wong & Clayton, 1985). A key feature of this model is the presence of extensive single-stranded DNA intermediates (Wanrooij et al., 2012).

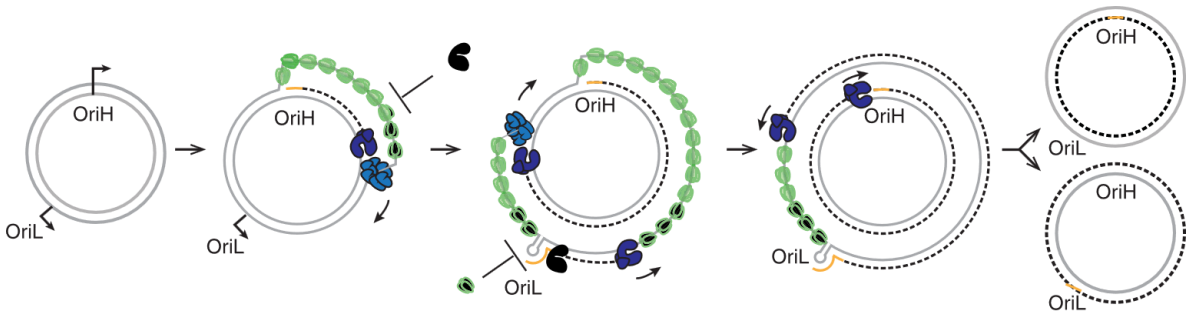


Figure 4. Strand-displacement model of mtDNA replication (schematic). Replication of the nascent H-strand initiates at OriH and proceeds unidirectionally, displacing the parental H-strand which is coated by mtSSB (green) to prevent unspecific primer formation and is unwound by the TWINKLE helicase (light blue) ahead of Pol γ (dark blue). When the fork exposes OriL, the parental H-strand folds into a stem-loop and POLRMT (purple) synthesises a short RNA primer (orange) from the poly-dT stretch in the loop, which Pol γ then extends to initiate continuous L-strand synthesis around the circle. Because the parental H-strand at OriL is already single-stranded, TWINKLE is not required for L-strand synthesis. The daughter molecule containing the nascent H-strand is initiated and terminated at OriH, while the complementary daughter molecule is initiated and terminated at OriL. (adapted from Falkenberg & Gustafsson, 2020).

However, subsequent analyses of replication intermediates using two-dimensional agarose gel electrophoresis (2D-AGE) challenged key aspects of the classic SDM, leading to the proposal of the Ribonucleotide Incorporation Throughout the Lagging Strand (RITOLS) model (Figure 5A) (Yasukawa et al., 2006). This model represents a paradigm shift, suggesting that the displaced parental H-strand is not single-stranded but is instead covered by pre-formed, processed RNA transcripts that hybridise to the DNA template. This creates a provisional lagging strand made of RNA, making the replication intermediate appear almost entirely duplex. According to this 'bootlace' mechanism, the incorporated RNA is later removed and replaced by DNA in a maturation step. While initiation remains asynchronous as in the SDM, the RITOLS model accounts for the observed duplex nature and RNase H sensitivity of replication intermediates found in highly purified mitochondria (Pohjoismäki et al., 2010; Reyes et al., 2013; Yang et al., 2002; Yasukawa et al., 2006).

A third mechanism, the strand-coupled model, has also been proposed, particularly in certain cell types or under conditions of mtDNA stress (Figure 5B) (Holt et al., 2000). This model posits a synchronous mode of replication that more closely resembles conventional, nuclear DNA replication. It involves the simultaneous synthesis of both the leading and lagging strands, with the lagging strand being synthesised discontinuously as Okazaki-like fragments. While less universally accepted as the primary replication mechanism under normal physiological conditions, evidence for this mode suggests it may serve as an alternative or backup pathway. The ongoing debate surrounding these models underscores the complexity and potential context-dependent flexibility of mtDNA replication in mammals. A deeper understanding of these processes is critical for elucidating the mechanisms of mitochondrial genome maintenance in both health and disease.

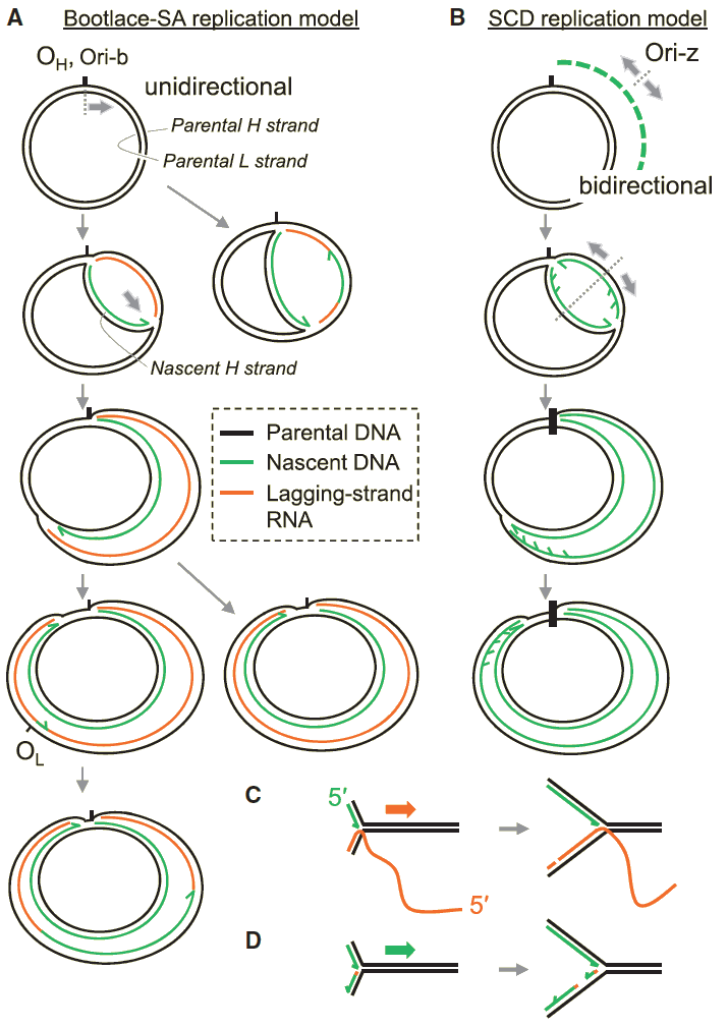


Figure 5. Proposed models of mtDNA replication. (A, C) Bootlace strand-asynchronous replication (Bootlace-SA / RITOLS model): This model illustrates unidirectional H-strand DNA synthesis (green line) initiating at Ori-b. Critically, the displaced parental H-strand is not left single-stranded but is hybridised with pre-formed RNA transcripts (orange lines), creating a provisional RNA lagging strand in a 'bootlace' fashion. This RNA is later replaced by DNA. (B, D) Strand-coupled DNA synthesis (SCD model): This model depicts bidirectional replication from a broad origin zone (Ori-z). Both leading and lagging strands are synthesised concurrently and are composed of DNA (green lines), similar to conventional nuclear replication. The lagging strand is formed from Okazaki-like DNA fragments. (Adapted from Yasukawa & Kang, 2018).

4.4. Replication initiation and control

In mammalian mitochondria, initiation of heavy-strand DNA replication is tightly connected to transcription from the light-strand promoter (LSP), such that transcriptional products generated immediately downstream of LSP form RNA/DNA hybrids (R-loops) that seed primer formation for DNA synthesis (Chang et al., 1985a; Chang & Clayton, 1985b; Gillum & Clayton, 1979). The mechanism of primer formation at OriH is a multi-step, regulated process. A significant fraction of transcription events initiated by POLRMT at LSP is prematurely terminated within a region containing three evolutionarily conserved sequence blocks (CSB1, CSB2 and CSB3) (Chang et al., 1985a; Kang et al., 1997; Pham et al., 2006; Posse et al., 2019). This termination event is not random; it is directed by specific sequence features, most notably the G-rich nature of CSB2 (Wanrooij et al., 2012; Wanrooij et al., 2010). During transcription of this region, the nascent RNA co-transcriptionally forms a stable hybrid G-quadruplex (G4) structure with the G-rich non-template DNA strand. This RNA/DNA hybrid structure serves to anchor the nascent RNA to the mtDNA, creating a persistent three-stranded structure known as an R-loop (Wanrooij et al., 2012; Pham et al., 2006; Xu & Clayton, 1995; Xu & Clayton, 1996; Sun et al., 2013). This R-loop is further stabilised by mtSSB, which likely binds to the displaced DNA strand and prevents re-annealing (Figure 6) (Sun et al., 2013).

However, the 3′-end of the RNA transcript within this newly formed R-loop is not directly accessible or competent to prime DNA synthesis by Pol γ (Posse et al., 2019). To become a functional primer, the R-loop must first undergo enzymatic processing. This critical step is performed by Ribonuclease H1 (RNase H1), an enzyme that specifically degrades the RNA strand of RNA/DNA hybrids (Cerritelli et al., 2003; Misic et al., 2022; Posse et al., 2019). RNase H1 cleaves the nascent RNA at multiple sites throughout the CSB2-CSB3 region, generating free 3′-OH ends. These newly created ends are then recognised and extended by Pol γ, marking the definitive transition from RNA priming to the initiation of H-strand DNA synthesis (Posse et al., 2019).

The balance between primer formation and the production of full-length, gene-coding transcripts from LSP is a key regulatory point. The mitochondrial transcription elongation factor (TEFM) has been shown to play a role in this switch. By binding to the elongating POLRMT, TEFM reduces the probability of premature termination at CSB2, thereby promoting processive, genome-length transcription (Agaronyan et al., 2015; Posse et al., 2015). The existence of this regulatory layer underscores the complexity of coordinating mtDNA replication with the expression of L-strand genes, both of which were long thought to originate from the single LSP promoter (Agaronyan et al., 2015; Chang & Clayton, 1985b; Pham et al., 2006; Wanrooij et al.,

2010). The recent discovery of a second light-strand promoter, LSP2, located downstream of OriH, has challenged this single-origin model, suggesting that L-strand gene expression and replication priming might be spatially and functionally decoupled, allowing for more flexible regulation of mitochondrial gene expression (Tan et al., 2022).

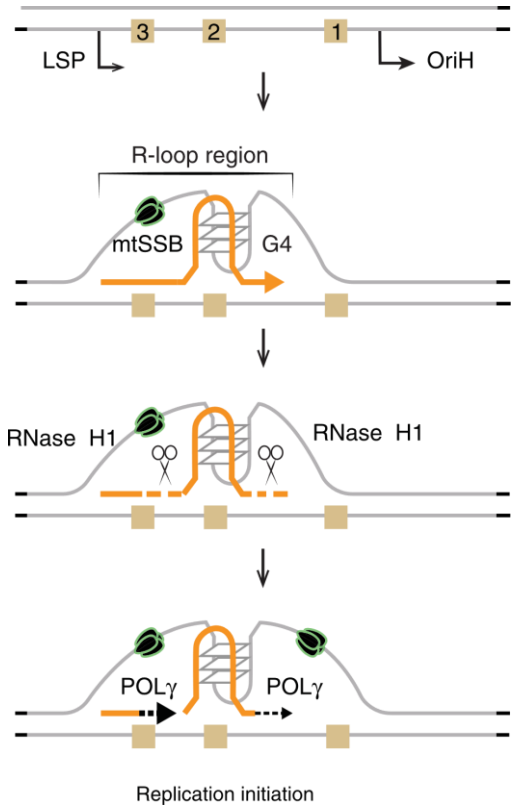


Figure 6. A model for transcription-coupled initiation of heavy-strand DNA replication at OriH. The process begins with transcription from the light-strand promoter (LSP) by POLRMT. At the conserved sequence block 2 (CSB2), a stable R-loop is formed, anchored by a hybrid G-quadruplex structure, leading to premature transcription termination. The nascent RNA is not a direct primer and requires processing by RNase H1, which generates free 3′-OH ends competent for extension by DNA polymerase γ. The entire structure is stabilised by mtSSB. (Adapted from Falkenberg & Gustafsson, 2020).

Beyond this initial regulation that determines whether transcription is extended or used for priming, replication encounters a second level of control once synthesis has begun. This checkpoint decides whether nascent heavy-strand synthesis will continue into full genome replication or terminate within the short displacement-loop (D-loop) region (Figure 1).

The majority of replication events initiated at OriH are prematurely terminated after synthesising only ~650 nucleotides, at a site known as the termination-associated sequence (TAS) (Bogenhagen & Clayton, 1978; Doda et al., 1981; Jemt et al., 2015). The resulting short DNA product, termed 7S DNA, remains hybridised to the template L-strand, forming a characteristic triple-stranded structure called the D-loop (Brown et al., 1986). This structure is subject to continuous metabolic turnover and is now viewed as a critical regulatory switch that controls the balance between abortive replication (the production of 7S DNA) and productive, full-genome replication.

The molecular mechanisms governing this regulation are not fully understood, but evidence points to the involvement of specific sequences and proteins. Two conserved, 15-nt palindromic sequence motifs ($ATGN_9CAT$) have been identified on each side of the D-loop: one within Conserved Sequence Block 1 (CSB1) and the other in the TAS region (Jemt et al., 2015). Although their precise function remains unknown, palindromic sequences often serve as binding sites for specific regulatory proteins. This hypothesis is supported by footprinting studies, which have demonstrated protected areas at both the CSB1 and TAS regions (McShane et al., 2024).

Furthermore, chromatin immunoprecipitation sequencing (ChIP-seq) experiments have revealed that Pol γ is enriched at both the 5′ and 3′ ends of the D-loop. This led to a model in which Pol γ remains associated with the 3′ end of the 7S DNA molecule. The reloading of the TWINKLE helicase at this site is proposed to enable the rapid re-initiation of full-length mtDNA synthesis, using the 7S DNA as a primer (Jemt et al., 2015). This model is consistent with observations that TWINKLE levels correlate with mtDNA copy number, and that conditions favouring increased replication (e.g., mild mtDNA depletion) lead to a decrease in 7S DNA levels accompanied by a concurrent increase in TWINKLE occupancy at TAS (Ikeda et al., 2015; Tyynismaa et al., 2004). Interestingly, reduced activity of POLRMT or RNase H1, which impairs de novo initiation, also leads to a rapid decrease in 7S DNA levels, suggesting that the cell compensates by favouring elongation from existing 7S DNA molecules (Kuhl et al., 2016; Misic et al., 2022).

Taken together, transcription coupled initiation at OriH produces R-loop derived precursors whose maturation depends on RNase H1. This critical dependency on RNase H1 for the first step of mtDNA replication underscores its central importance. To fully appreciate its function, it is necessary to examine the enzyme specific molecular architecture, catalytic mechanism, and the clinical consequences of its dysfunction.

RNase H1 is an essential endonuclease that belongs to the ancient Retroviral Integrase Superfamily (RISF) and is defined by its specific ability to hydrolyse the RNA strand within RNA/DNA hybrid structures. The RNase H family is broadly classified into distinct types, with

mammals possessing two major classes: RNase H1 and RNase H2, while a third class, RNase H3, is found in some bacteria and archaea (Nowotny, 2009). Although both mammalian enzymes resolve RNA/DNA hybrids, they have distinct substrate preferences and cellular roles; RNase H2 is uniquely able to excise single ribonucleotides embedded within a DNA duplex (Rydberg & Game, 2002; Qiu et al., 1999), whereas RNase H1 is the only isoform present in mitochondria, making it central to mtDNA metabolism (Cazenave et al., 1994; Cerritelli & Crouch, 1998; Cerritelli & Crouch, 2009). In mammalian cells, the *RNASEH1* gene produces dually localised isoforms for the nucleus and mitochondria via alternative translation initiation sites, with the mitochondrial form containing an N-terminal Mitochondrial Targeting Sequence (MTS) (Liang et al., 2017; Suzuki et al., 2010). Its indispensable role in mtDNA maintenance is demonstrated by the embryonic lethality and catastrophic mtDNA depletion observed in knockout mice (Cerritelli et al., 2003). Consistent with its essential nature, pathogenic mutations in the human *RNASEH1* gene are now known to cause adult-onset mitochondrial encephalomyopathy, a severe disorder characterised by symptoms such as chronic progressive external ophthalmoplegia (PEO), muscle weakness, cerebellar atrophy, and the presence of ragged-red and COX-negative fibres in muscle biopsies (Reyes et al., 2015).

The enzyme exhibits a modular architecture, consisting of an N-terminal Hybrid-Binding Domain (HBD), a flexible linker, and a C-terminal catalytic domain (CTD) (Figure 7) (Cerritelli & Crouch, 2009). The catalytic core adopts the characteristic RNase H fold, centred on a mixed five-stranded β-sheet, and contains a conserved DEDD motif that coordinates two divalent metal ions (typically Mg^{2+}) essential for catalysis. The catalytic mechanism follows a canonical two-metal-ion model, where one metal ion activates a nucleophilic water molecule for an SN2-type attack on the scissile phosphate, while the second metal ion stabilises the transition state and the leaving group (Nowotny et al., 2007; Yang et al., 2006).

The HBD is critical for the enzyme's function, conferring high substrate affinity and processivity, which allows for the efficient degradation of long RNA tracts (Gaidamakov et al., 2005; Nowotny et al., 2008). Structurally, the HBD binds within the minor groove of the hybrid, showing a more than 25-fold preference for RNA/DNA hybrids over dsRNA. It uses an aromatic patch (W43, F58) to recognise the deoxyribose backbone of the DNA strand and a conserved loop to interact with the 2'-OH groups of the RNA strand. This geometry creates steric hindrance for a second RNA strand, thus ensuring specificity against dsRNA (Nowotny et al., 2008).

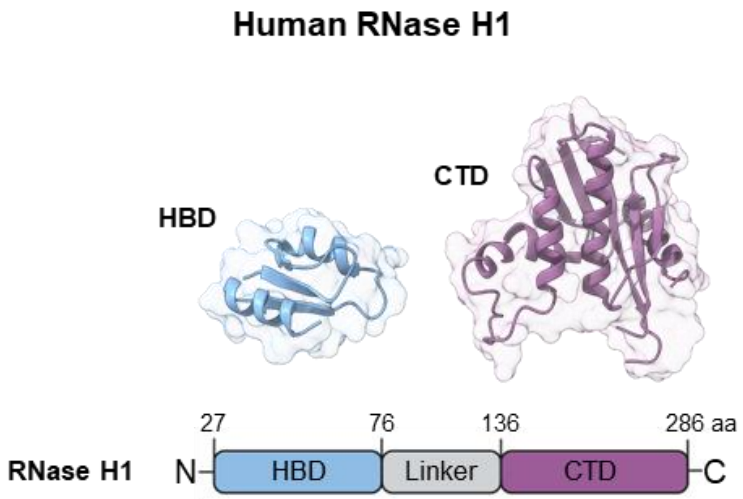


Figure 7. Domain organisation of human RNase H1. Structure of the hybrid-binding domain of holo RNase H1 (blue; PDB: 3BSU (Nowotny et al., 2008)) and its catalytic domain (purple; PDB: 2QKB (Nowotny et al., 2007)). Structure was rendered in UCSF ChimeraX (Pettersen et al., 2021). Below: Schematic representation of the RNase H1 domain structure with marked domain boundaries.

The functional importance of these domains is underscored by the location of pathogenic variants. Most disease-causing mutations, such as p.Val142Ile and p.Ala185V, are found within the catalytic domain and severely compromise the enzyme's structural stability, leading to a near-unfolded state or a significantly reduced melting temperature (Al-Behadili et al., 2018). This instability results in drastically reduced enzymatic activity *in vitro* and impaired mtDNA replication and primer retention *in vivo* (Suzuki et al., 2010). Conversely, mutations have also been identified in the hybrid-binding domain (HBD), such as the p.Tyr29Cys variant, which leads to a paradoxically increased nuclease activity. However, this hyperactivity is coupled with an impaired primer formation, demonstrating that the HBD is crucial not only for substrate binding but also for the precise regulation of catalysis required for correct replication initiation (Misic et al., 2022).

While RNase H1 is indispensable for creating a functional primer at the start of replication, its fundamental ability to process RNA/DNA hybrids makes it equally critical at the end of the process: the removal of these primers to allow for the final ligation of the daughter strands.

4.5. RNA Primer Removal and Termination

Upon completion of mtDNA synthesis, replication must be properly terminated to yield intact, covalently closed circular daughter molecules. Termination entails the removal of RNA primers from the nascent DNA strands used to initiate DNA synthesis at both origins, the filling of any

remaining gaps, and the ligation of nascent strand ends (Macao et al., 2015; Uhler & Falkenberg, 2015). Because initiation at OriH and OriL occurs asynchronously and in distinct genomic contexts, the requirements and molecular mechanisms governing primer removal differ markedly between the two origins. These residual ribonucleotides are incompatible with efficient ligation by DNA ligase III, necessitating the action of a second nuclease activity (Al-Behadili et al., 2018).

At OriL, POLRMT synthesises a short RNA primer of approximately twenty to twenty-five nucleotides from the poly-dT stretch in the stem-loop structure exposed when the replication fork passes this origin (Fusté et al., 2010). Following initiation of L-strand synthesis by Pol γ, the RNA primer must be excised before ligation can seal the nascent strand. RNase H1 is essential for this process (Cerritelli et al., 2003; Holmes et al., 2015; Ruhanen et al., 2011), cleaving most of the primer but leaving one to three ribonucleotides at the RNA-DNA junction owing to its requirement for four consecutive ribonucleotides as substrate (Nowotny et al., 2007).

Since RNase H1 alone does not fully resolve the OriL primer, its function should be considered within the broader cooperation of mitochondrial replication proteins. Beyond its direct cleavage activity, RNase H1 likely operates as part of a multi-protein replication module that synchronises primer removal with ongoing DNA synthesis and ligation. Biochemical and cellular data demonstrate physical and functional links between RNase H1, the Pol γ holoenzyme, and DNA ligase III (Karlowicz et al., 2022), supporting a model in which processing of the RNA-DNA junction is coordinated within the replication machinery. Complementary studies in other organisms indicate that single-stranded DNA binding proteins can stabilise RNA/DNA hybrids and enhance RNase H1 processivity, as shown for yeast RPA (Li et al., 2024) and for mtSSB in Drosophila (González de Cózar et al., 2020). Together, these findings support a view in which RNase H1 activity is modulated through protein-protein interactions that promote more efficient primer removal and ligation, thereby ensuring continuity of the nascent strand. Such cooperation may enhance RNase H1 processivity and facilitate near-complete primer processing; however, current evidence indicates that an additional nuclease is still required to remove the remaining ribonucleotides, as described in the following models (Figure 8).

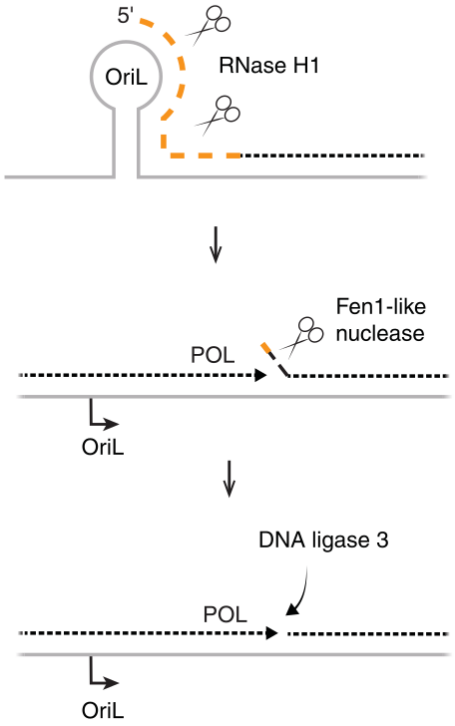


Figure 8. Processing of the RNA Primer at OriL Primer removal at the OriL is a sequential process. Initially, RNase H1 cleaves the ~25 nucleotide RNA primer, leaving a short 1-3 ribonucleotide fragment at the 5′ end of the new L-strand. As DNA polymerase γ completes DNA synthesis, it displaces the 5′ end, creating a small flap. This flap is then proposed to be excised by a FEN1-like nuclease, which generates a ligatable nick that is ultimately sealed by DNA Ligase III to ensure strand continuity. (Adapted from Falkenberg & Gustafsson, 2020).

Two models have been proposed to account for the removal of these terminal ribonucleotides: a flap-dependent pathway and a flap-independent pathway (Al-Behadili et al., 2018; Wu et al., 2019; Karlowicz et al., 2022). In the flap-dependent model, Pol γ continues synthesis beyond the primer-template junction, displacing the residual ribonucleotides into a short single-stranded flap that can subsequently be cleaved by a flap endonuclease (Macao et al., 2015; Uhler & Falkenberg, 2015). Although flap structure-specific endonuclease 1 (FEN1) has been suggested as a candidate enzyme, its mitochondrial localisation and functional significance remain controversial, and depletion studies have failed to demonstrate a clear requirement for FEN1 in mitochondrial primer removal (Akbari et al., 2008; Szczesny et al., 2008; Tann et al., 2011). Alternative nucleases, including DNA2 and MGME1, have been implicated in flap processing, yet neither appears wholly sufficient to resolve short RNA-containing flaps at OriL (Al-Behadili et al., 2018; Szczesny et al., 2013; Uhler et al., 2016).

The flap-independent pathway, by contrast, proposes direct exonucleolytic removal of the residual ribonucleotides without requiring strand displacement. In this scenario, a nuclease with

the capacity to excise RNA dinucleotides from the 5′ end of DNA would generate a ligatable substrate in a single catalytic step. Recent biochemical and structural studies have demonstrated that human mitochondrial 5′-exonuclease G, EXOG, possesses precisely this activity: it functions as a 5′ exonuclease that preferentially cleaves at RNA-DNA junctions, excising dinucleotides and thereby removing the RNase H1-resistant ribonucleotides without the need for flap formation (Karlowicz et al., 2022; Wu et al., 2019).

EXOG localises exclusively to mitochondria (Cymerman et al., 2008) and plays a fundamental role in maintaining mtDNA integrity through its involvement in base excision repair (BER) and RNA primer removal (Karlowicz et al., 2022; Szymanski et al., 2017; Wu et al., 2019). It is a member of the His-Me finger nuclease superfamily and a paralog of the apoptotic endonuclease EndoG (Wu et al., 2020; Yang, 2011). EXOG functions as a symmetrical homodimer in its unbound state, but becomes functionally asymmetrical upon substrate binding, exhibiting half-site reactivity (Szymanski et al., 2017). The defining structural feature of EXOG is a unique C-terminal Wing domain that is absent in EndoG (Figure 9). This domain is crucial, as it converts the enzyme's latent non-specific endonuclease activity into a highly specific 5′-exonuclease function (Szymanski et al., 2017).

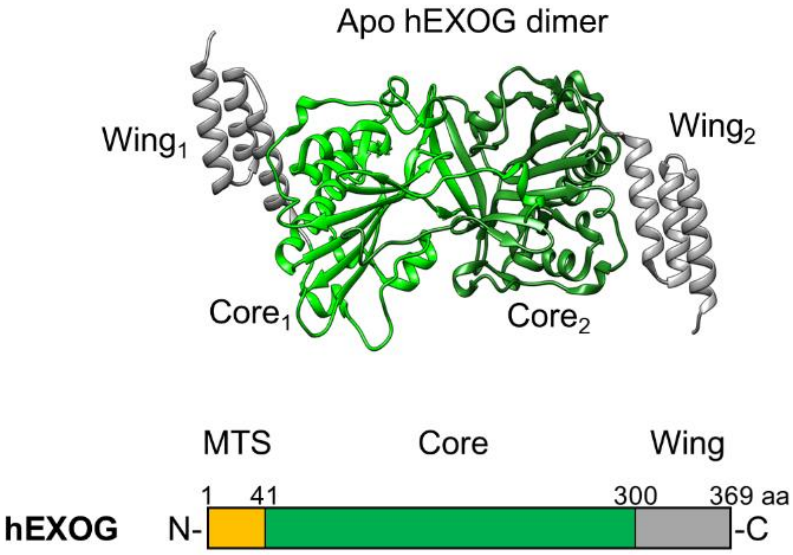


Figure 9. **Domain organisation of human EXOG**. Structure of the apo hEXOG-Δ58 dimer (PDB: 5T40 (Szymanski et al., 2017)), which lacks the mitochondrial targeting sequence, transmembrane segment, and the predicted intrinsically disordered region. Below: Schematic representation of the domain organisation of human EXOG with indicated domain boundaries. (Adapted from Karlowicz et. al, 2025).

The Wing domain acts as a ‘tape-measure’, precisely positioning the duplex DNA substrate to ensure cleavage occurs between the second and third nucleotides from the 5′ end, thereby excising a dinucleotide product. This activity is critically important for mitochondrial BER, as the resulting two-nucleotide gap is the optimal substrate for gap-filling synthesis by Pol γ. Upon DNA binding, EXOG induces a B-to-A form conformational change in DNA, which facilitates cleavage (Szymanski et al., 2017).

Importantly, reconstitution experiments with purified mitochondrial proteins have shown that EXOG, in conjunction with RNase H1 and DNA ligase III, can support complete primer removal and ligation of both linear and circular OriL-like substrates *in vitro*. The physical interaction between EXOG and RNase H1, confirmed by both microscale thermophoresis and proximity ligation assays, further supports a model in which these two-nucleases act coordinately in a two-nuclease pathway for flap-independent RNA primer processing (Karlowicz et al., 2022). Nonetheless, definitive *in vivo* evidence for EXOG role at OriL remains incomplete, and it is plausible that both flap-dependent and flap-independent mechanisms operate in parallel, with their relative contributions modulated by local DNA topology, the cellular context, and the availability of specific nucleases.

Termination of H-strand synthesis at OriH presents a mechanistically more intricate challenge than at OriL, necessitating an extensive post-synthetic maturation of the nascent strand (Figure 10). This process is defined by a notable spatial discrepancy: while RNA-to-DNA transitions primarily occur within the conserved sequence block (CSB) region, particularly near CSB2 (Posse et al., 2019), the mature 5′ end of the nascent H-strand is located approximately 100 nucleotides downstream at position 191 (Clayton, 1991; Nicholls & Minczuk, 2014). This implies an extensive processing pathway where both the RNA primer and a segment of the nascent DNA are removed (Jemt et al., 2015).

This maturation is believed to be a two-nuclease event, initiated by RNase H1, which removes the bulk of the RNA primer (Misic et al., 2022), followed by the action of the mitochondrial genome maintenance exonuclease 1 (MGME1) (Uhler et al., 2016). For MGME1 to function, the nascent H-strand must first be displaced from its template, forming a single-stranded flap substrate. The precise mechanism of this displacement remains unclear, though it has been suggested that transcription-coupled activity of POLRMT could be responsible (Falkenberg & Gustafsson, 2020). MGME1, a 5′-3′ exonuclease, then processes this chimeric flap, but its cleavage is imprecise, generating a heterogeneous pool of nicks, gaps, and short residual flaps (Uhler et al., 2016).

Crucially, these products are often not directly ligatable. The process requires the intervention of Pol γ, which utilises its polymerase and 3′-5′ exonuclease activities to 'edit' the imprecise ends, creating a perfectly aligned nick that can be sealed by DNA ligase III (Karlowicz et al., 2022). While this RNase H1-MGME1 axis is considered the primary pathway, the unique ability of exonuclease EXOG to cleave directly at RNA-DNA junctions suggests a potential complementary, flap-independent mechanism might also contribute to primer removal at OriH (Karlowicz et al., 2022; Wong et al., 2008).

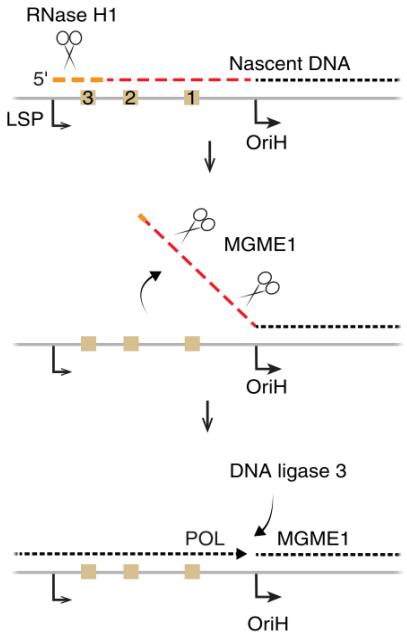


Figure 10. Maturation of the Nascent H-Strand at OriH. The 5′ terminus of the nascent heavy-strand undergoes extensive processing at its origin (OriH), which relocates the strand end from the CSB3/CSB2 region to position 191. This maturation event requires the excision of both the RNA primer and an adjacent segment of approximately 100 nucleotides of newly synthesised DNA. This process is a general feature of 5′ end formation, as it also occurs on the short 7S DNA molecule. The removal begins with RNase H1 degrading the RNA primer, leaving a 1-3 ribonucleotide remnant. The subsequent excision of these final ribonucleotides and the downstream DNA segment is thought to be performed by a nuclease acting on a single-stranded DNA flap. While the mitochondrial nuclease MGME1 is a strong candidate for this function, it remains possible that other nucleases could also contribute to this processing step. A critical but unresolved aspect is the mechanism by which this essential flap substrate is generated. (Adapted from Falkenberg & Gustafsson, 2020).

Failure in any step of this intricate processing, particularly due to defects in MGME1 or the exonuclease domain of Pol γ, prevents successful ligation. This results in a persistent nick in the H-strand, which, during the subsequent round of replication, can lead to a double-strand break and the formation of characteristic linear mtDNA fragments spanning the major arc. The clinical repercussions of such catastrophic events are profound, as these molecular defects are the direct

cause of a severe class of human mitochondrial diseases known as mtDNA maintenance disorders (Alexeyev et al., 2013; Gustafson et al., 2020; Gustafsson et al., 2016; Zhao & Sumberaz, 2020). The integrity of the mitochondrial genome is non-negotiable for cellular health, as it encodes 13 essential proteins for the oxidative phosphorylation (OXPHOS) system responsible for generating ATP Consequently, the accumulation of structurally compromised mtDNA, whether in the form of linear fragments, large-scale deletions or overall depletion, directly impairs cellular energy production, a failure that manifests most acutely in high-energy-demand tissues such as the brain, skeletal muscle, heart and liver (Greaves et al., 2012; Ylikallio & Suomalainen, 2012). This molecular cascade, from a single unresolved nick to systemic bioenergetic collapse, forms the central pathogenic mechanism linking enzymatic defects in factors like RNase H1, FEN1, DNA2, and Lig III to observable disease (Uhler & Falkenberg, 2015). For example, the embryonic lethality observed in mice lacking RNase H1 underscores its indispensable role, as mtDNA replication ceases, leading to a catastrophic drop in copy number and developmental arrest (Cerritelli et al., 2003).

This genomic instability is the direct aetiological basis for a wide spectrum of devastating and often progressive clinical syndromes. Pathogenic mutations in the nuclear genes encoding the core mitochondrial replisome, including *POLG, TWNK, RNASEH1*, and *MGME1*, are known to disrupt the fidelity of mtDNA replication and repair, leading to the accumulation of these molecular lesions (Alexeyev et al., 2013; Copeland, 2014; Greaves et al., 2012; Gustafson et al., 2020; Gustafsson et al., 2016; Ylikallio & Suomalainen, 2012; Zhao & Sumberaz, 2020). Clinically, this manifests as disorders such as adult-onset mitochondrial encephalomyopathy, which is characterised by chronic progressive external ophthalmoplegia (PEO), severe muscle weakness, exercise intolerance, and cerebellar signs like ataxia and dysarthria. In more severe cases, these defects present in infancy or childhood as the often-fatal Alpers' syndrome, Pearson syndrome, or Kearns-Sayre syndrome (KSS) (Moraes et al., 1995; Schon et al., 1989). A central event preceding these outcomes is replication stalling, a universal step where the replication fork arrests, rendering the exposed single-stranded DNA vulnerable to breakage and illicit recombination, which in turn generates circular deletions. The clinical presentation is frequently governed by a 'threshold effect', whereby symptoms only emerge once the burden of deleted or damaged mtDNA surpasses a critical level (often 60-80%) within a given tissue (Fontana & Gahlon, 2020). Therefore, elucidating the precise enzymatic mechanisms that ensure faithful replication termination is not only a question of fundamental biology; it is essential for understanding the pathogenesis of these currently incurable diseases and for providing the molecular foundation needed to develop future therapeutic strategies aimed at preserving mitochondrial genome integrity

5. Aims of the Thesis

The fidelity of mitochondrial DNA (mtDNA) replication depends on the complete and accurate removal of RNA primers to generate ligatable DNA ends. RNase H1 initiates primer degradation but cannot excise the final one to three ribonucleotides at the RNA-DNA junction, thereby necessitating a coordinated handover to partner enzymes. Among these, the 5′-exonuclease EXOG directly removes the remaining ribonucleotides, ensuring completion of primer processing. The replicative polymerase Pol γ, which physically interacts with RNase H1, represents another essential factor in this pathway, yet the mechanistic consequences of this association remain unknown. Although functional and physical links between RNase H1 and each of these partners have been demonstrated, the molecular details governing the respective interactions remain poorly understood. Critical knowledge gaps persist regarding the specific protein interfaces, the order and coordination of their sequential actions on shared RNA-DNA intermediates, and the structural organisation of the resulting complexes. This thesis addresses these questions by dissecting two distinct enzymatic collaborations, RNase H1 with EXOG and RNase H1 with Pol γ, to clarify their individual contributions to the terminal stages of mitochondrial primer maturation.

The objectives of this study are:

- To define the functional coordination between RNase H1 and its partners EXOG and Pol γ, by reconstituting primer-removal reactions *in vitro* and assessing how pathogenic RNase H1 variants affect EXOG-dependent processing, and by investigating the interplay between RNase H1 and Pol γ on defined gapped substrates.
- To elucidate the molecular interfaces that govern these interactions through an integrated computational-biochemical approach combining AlphaFold-based predictions with SPR analysis and targeted mutagenesis.
- To determine the three-dimensional structures of the key complexes involved in mitochondrial replication termination, employing X-ray crystallography for the RNase H1-EXOG complex and single-particle cryo-EM for the Pol γ-RNase H1-RNA/DNA substrate assembly.

6. Materials and methods

Throughout this thesis, Pol γ denotes the human mitochondrial DNA polymerase holoenzyme Pol γAB unless stated otherwise. Individual subunits are referred to as Pol γA and Pol γB. RNase H1 refers to the human full length wild-type enzyme unless indicated; in figure labels it is abbreviated as RH1 for clarity. Truncated constructs are referred to as CTD for the C-terminal catalytic domain, HBD for the N-terminal hybrid-binding domain, and CTD Linker or HBD Linker where applicable.

6.1. Protein overexpression

6.1.1. Human mitochondrial Ribonuclease H1 (RNase H1)

A plasmid encoding human RNase H1 (residues 27-286) in the pET-15b vector, containing an N-terminal His_6 tag followed by a thrombin cleavage site, was obtained from the laboratory of Professor Marcin Nowotny. The plasmid was transformed into *Escherichia coli* BL21 (DE3) pLysS cells. A pre-culture was prepared from a single bacterial colony and used to inoculate a larger culture in lysogeny broth (LB) medium supplemented with ampicillin. Cultures were grown at 37 °C with shaking until reaching an optical density at 600 nm (OD_{600}) of approximately 0.6. At this point, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), and the cultures were incubated overnight at 16 °C to promote soluble protein expression. Cells were harvested by centrifugation and stored at -80 °C until purification.

6.1.2. Human mitochondrial DNA polymerase γ catalytic subunit A (Pol γA)

Recombinant human Pol γA (the catalytic subunit of DNA polymerase γ), lacking the mitochondrial targeting sequence and carrying a truncated polyglutamine stretch at positions 43-52, was expressed in Spodoptera frugiperda (Sf9) insect cells using a baculovirus expression system based on the pBacPak9 vector (Clontech). Infected cell pellets were obtained from the EMBL Protein Expression and Purification Core Facility, and stored at -80 °C upon arrival.

6.1.3. Human mitochondrial DNA polymerase γ accessory subunit B (Pol γB)

Recombinant human Pol γB was expressed in *Escherichia coli* BL21 Rosetta 2 (DE3) pLysS cells (Novagen) transformed with a pET22b(+) vector encoding the full-length sequence of the accessory subunit. Bacterial cultures were grown in LB medium supplemented with ampicillin and chloramphenicol at 37 °C with shaking. When the OD_{600} reached approximately 0.55, protein expression was induced by the addition of 0.4 mM IPTG, and cultures were incubated for an

additional 7 hours at 25 °C. Cells were harvested by centrifugation and stored at -80 °C until purification.

6.1.4. Human mitochondrial 5′-exonuclease G (EXOG)

Recombinant human EXOG, lacking the mitochondrial targeting sequence (ΔN58) and carrying a C-terminal His_6 tag, was expressed in *Escherichia coli* Rosetta (DE3) cells transformed with a pET22b(+) expression vector. Bacterial cultures were grown in LB medium supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) at 37 °C with shaking. Upon reaching an OD_{600} of 0.6, protein expression was induced with 1 mM IPTG, and the cultures were further incubated for 3 hours at 30 °C. Cells were harvested by centrifugation, flash-frozen in liquid nitrogen, and stored at -80 °C until use.

6.2. Protein purification

6.2.1. RNase H1

Recombinant human RNase H1 (residues 27-286) was expressed in *Escherichia coli* and purified by sequential affinity, proteolytic, ion-exchange, and size-exclusion chromatography. All RNase H1 variants used in this study, including the wild-type protein, individual domain constructs (CTD and HBD), domain-linker combinations (CTD-Linker and HBD-Linker), as well as two disease-associated variants (A185V and V142I), were purified using the same modified procedure. Bacterial cultures were harvested by centrifugation and resuspended in RNase H1 lysis buffer composed of 20 mM HEPES pH 7.0, 1 M NaCl, 5% (v/v) glycerol, 2.8 mM β-mercaptoethanol, and 10 mM imidazole. The buffer was supplemented with protease inhibitors (E64, pepstatin A, and leupeptin). Cells were lysed by sonication, and the lysate was clarified by high-speed centrifugation. The supernatant was applied to a Ni^{2+}-NTA affinity column pre-equilibrated with the same lysis buffer. Unbound material was collected and reapplied to the resin to maximise binding efficiency. The column was washed with buffer containing 60 mM imidazole, and bound proteins were eluted with buffer containing 300 mM imidazole. Affinity-purified fractions were pooled and diluted with a no-salt buffer (20 mM HEPES pH 7.0, 5% (v/v) glycerol, 2.8 mM β-mercaptoethanol) to reduce ionic strength. The sample was then dialysed against thrombin cleavage buffer containing 20 mM HEPES pH 7.0, 100 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, and 0.5 mM EDTA. Thrombin digestion was performed overnight at 4 °C using a commercial thrombin preparation (Sigma-Aldrich), and the efficiency of tag removal was confirmed by SDS-PAGE. Unless stated otherwise, all SDS-PAGE analyses were performed using PageRuler™ Prestained Plus Protein Ladder (Thermo Fisher Scientific) as the molecular weight marker. The cleaved protein was subjected to cation exchange chromatography using a HiTrap SP HP column (Cytiva) equilibrated in RNase H1 buffer SPA (20 mM HEPES pH 7.0,

100 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, 0.5 mM EDTA). After washing, bound proteins were eluted in a stepwise fashion using high-salt buffer (identical composition to RNase H1 buffer SPA, but 1 M NaCl). Fractions corresponding to the elution peak were analysed by SDS-PAGE and pooled. Final purification was performed by size-exclusion chromatography on a Superdex™ 200 column (Cytiva) equilibrated in 20 mM HEPES pH 7.0, 200 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, and 0.5 mM EDTA. Elution profiles were monitored spectrophotometrically, and fractions corresponding to the main peak were collected. The final sample was dialysed against the same buffer supplemented with 30% glycerol, concentrated, aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C. Protein concentration was determined spectrophotometrically using an extinction coefficient of $\varepsilon_{280} = 47\,565\ M^{-1}cm^{-1}$

6.2.2. Pol γA

Sf9 cell pellets were thawed on ice and resuspended in Pol γA lysis buffer composed of 15 mM HEPES pH 7.6, 250 mM sucrose, 75 mM KCl, 2.25 mM $CaCl_2$, 1.5 mM magnesium acetate, 1.25% (v/v) glycerol, 0.375% (v/v) NP-40, and 0.025% (v/v) Triton X-100. The buffer was supplemented with protease inhibitors (pepstatin A, E-64, leupeptin, and PMSF) immediately before use. Cell lysis was performed by gentle mechanical agitation at 4 °C. After clarification by low-speed centrifugation, the supernatant was adjusted to a final concentration of 0.5 M KCl by dropwise addition of 3 M KCl while stirring on ice. The solution was incubated briefly and centrifuged at high speed to remove precipitated material. The resulting clarified extract was loaded onto gravity flow affinity columns packed with cobalt-based Talon Superflow resin (Cytiva) pre-equilibrated in wash buffer containing 20 mM HEPES pH 7.5, 200 mM KCl, 5% (v/v) glycerol, and 5 mM imidazole. After sample application, the columns were washed with the same buffer to remove unbound proteins. Elution was carried out using buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5% (v/v) glycerol, and 200 mM imidazole. Elution fractions were analysed by SDS-PAGE, pooled, and concentrated using centrifugal filtration (50 kDa MWCO). Final purification was performed by size-exclusion chromatography on a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare), equilibrated in 20 mM HEPES pH 7.5, 140 mM KCl, 5% (v/v) glycerol, 5 mM β-mercaptoethanol, and 0.1 mM EDTA. Elution profiles were monitored spectrophotometrically, and fractions corresponding to the main peak were analysed by SDS-PAGE. Selected fractions were pooled, concentrated, aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C until use. Protein concentration was determined spectrophotometrically using an extinction coefficient of $\varepsilon_{280} = 242\,295\ M^{-1}cm^{-1}$.

6.2.3. Pol γB

Frozen bacterial pellets were thawed on ice and resuspended in Pol γB lysis buffer containing 20 mM HEPES pH 8.0, 300 mM KCl, 5% (v/v) glycerol, and (v/v) 0.1% Triton X-100, supplemented with protease inhibitors (pepstatin A, E-64, and leupeptin). Cells were lysed by sonication, and the lysate was clarified by centrifugation. Nucleic acids were removed by the addition of polyethyleneimine to a final concentration of 0.1%, followed by incubation and further centrifugation. The supernatant was dialysed against buffer containing 20 mM HEPES pH 8.0, 200 mM KCl, 5% (v/v) glycerol, 10 mM imidazole, and 1 mM β-mercaptoethanol using multiple buffer exchanges at 4 °C. The dialysed lysate was loaded onto a Ni^{2+}-NTA affinity column pre-equilibrated in the same buffer. After washing, bound proteins were eluted with buffer containing 20 mM HEPES pH 8.0, 100 mM KCl, 5% (v/v) glycerol, 200 mM imidazole, and 1 mM β-mercaptoethanol. Eluted fractions were pooled, diluted to reduce salt concentration, and subjected to cation exchange chromatography using a HiTrap SP HP column (Cytiva) equilibrated in buffer A (20 mM HEPES pH 8.0, 60 mM KCl, 5% (v/v) glycerol). The protein was eluted using a linear gradient of buffer B (20 mM HEPES pH 8.0, 1 M KCl, 5% (v/v) glycerol), and fractions containing Pol γB were identified by SDS-PAGE, pooled, and concentrated. Final purification was performed by preparative size-exclusion chromatography using a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare) equilibrated in 20 mM HEPES pH 8.0, 140 mM KCl, 5% (v/v) glycerol, and 5 mM β-mercaptoethanol. The oligomeric state and purity of the protein were confirmed by elution profile and SDS-PAGE analysis. Protein concentration was determined spectrophotometrically using an extinction coefficient of $\varepsilon_{280} = 143\ 280\ M^{-1}cm^{-1}$ for the dimer.

6.2.4. Pol γAB holoenzyme

To assemble the Pol γAB holoenzyme complex, purified Pol γA and dimeric Pol γB were mixed at a 1:1.1 ratio based on A280. The mixture was concentrated using a 100 kDa molecular weight cutoff centrifugal filters (Vivaspin, Sartorius) and applied to a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare) equilibrated in 20 mM HEPES pH 7.5, 140 mM KCl, 5 mM β-mercaptoethanol, and 2% (v/v) glycerol. Pol γAB-containing fractions were pooled and further concentrated using a 100 kDa cutoff centrifugal device. The extinction coefficient used for the complex was $\varepsilon_{280} = 385\ 575\ M^{-1}cm^{-1}$.

6.2.5. EXOG

Frozen bacterial pellets were resuspended in denaturing EXOG lysis buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 0.1% CHAPS, 10 mM DTT, 10% sucrose, 10 mM imidazole), supplemented with protease inhibitors (pepstatin A, E-64, leupeptin). Lysozyme was

added, and the suspension was incubated on a magnetic stirrer at 4 °C. Cells were lysed by sonication, and insoluble material was removed by high-speed centrifugation. The remaining pellet was resuspended in solubilisation buffer (10 mM Tris-HCl pH 8.2, 8 M urea, 20 mM β-mercaptoethanol, 200 mM NaCl, 10 mM imidazole) and incubated with stirring at 4 °C overnight to allow protein solubilisation. The solubilised protein fraction was clarified by centrifugation and loaded onto a Ni^{2+}-NTA agarose column pre-equilibrated in the same buffer. The resin was sequentially washed with denaturing wash buffers containing decreasing concentrations of urea (6 M and 3.5 M), followed by elution with 3.5 M urea buffer containing 200 mM imidazole. Eluted fractions were pooled, aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C. Selected fractions were thawed and gradually refolded by stepwise dialysis. First, the pooled eluate was diluted in buffer containing 3.5 M urea and dialysed against refolding buffer (50 mM HEPES pH 7.5, 0.5 M urea, 1.5 M sorbitol, 1 mM TCEP, 24 mM NaCl, 1 mM KCl) at room temperature. This was followed by overnight dialysis at 4 °C against 300 mM NaCl buffer. The refolded protein was clarified by centrifugation and concentrated by ultrafiltration (30 kDa MWCO). Final purification was carried out by size-exclusion chromatography on a HiLoad™ 16/600 Superdex™ 200 pg column (GE Healthcare), equilibrated in 20 mM HEPES pH 8.0, 300 mM KCl, 1 mM TCEP, and 5% (v/v) glycerol. Fractions corresponding to the dimeric EXOG peak were identified by SDS-PAGE, pooled, concentrated, and flash-frozen in aliquots. Protein concentration was determined spectrophotometrically using an extinction coefficient of $\varepsilon_{280} = 77\,030\ M^{-1}cm^{-1}$ for the dimer.

6.3. Circular dichroism (CD) spectroscopy

Circular dichroism spectra were recorded on a Jasco J-1500 spectropolarimeter at 20 °C using a 1 mm path length quartz cuvette and CD buffer (20 mM potassium phosphate pH 8.0, 140 mM KCl). Spectra were collected from 190 to 260 nm with a 1 nm step size at a protein concentration of 2.5 µM. The spectral bandwidth was set to 1 nm, with a scan speed of 50 nm/min and a data pitch of 0.1 nm. Each spectrum represents the average of three consecutive scans. All spectra were baseline-corrected by subtraction of the buffer spectrum recorded under identical conditions. Protein concentrations were determined spectrophotometrically using specific extinction coefficients prior to CD measurement. For thermal denaturation experiments, changes in ellipticity at 208 nm were monitored during a continuous temperature ramp from 20 °C to 97 °C at a rate of 0.5 °C/min using a Peltier temperature control system. Samples were equilibrated at the starting temperature for at least 5 minutes prior to measurement. Melting temperatures (T_m) were determined using Spectra Manager v2.0 (JASCO) as the inflection point

of the unfolding curve. These measurements were performed with the support of Dr hab. Rafał Dutkiewicz, prof. UG.

6.4. Thermal unfolding

Thermal stability was assessed using a Tycho NT.6 instrument (NanoTemper Technologies), which monitors changes in intrinsic protein fluorescence during thermal unfolding in real time. This capillary-based technique measures fluorescence emission at 330 and 350 nm, wavelengths corresponding primarily to the emission spectra of tyrosine and tryptophan residues, respectively. The unfolding profile was obtained by recording the fluorescence ratio (350 nm/330 nm) as a function of increasing temperature. Protein samples were loaded into standard Tycho capillaries (NanoTemper). A temperature gradient from 35 °C to 95 °C was applied at a constant heating rate of 30 °C/min. The inflection temperature (T_i), indicating the midpoint of the unfolding transition, was automatically determined by the Tycho evaluation software as the maximum of the first derivative of the fluorescence ratio curve.

6.5. Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed on a Biacore T200 system (Cytiva) at 25 °C to investigate protein-protein interactions. CM5 sensor chips (Cytiva) were functionalised using standard amine coupling chemistry. Prior to immobilisation, optimal conditions were identified by pH scouting in sodium acetate or phosphate buffers (pH 3.5-8.0), selecting conditions that provided efficient coupling without compromising baseline stability. Based on these tests, recombinant Pol γA was diluted to 20 ng/µL in 10 mM sodium phosphate buffer pH 6.5 and immobilised to a surface density of approximately 971 response units (RU). Pol γB was diluted to 5 ng/µL in 10 mM sodium acetate buffer pH 5.5 and immobilised to approximately 1178 RU, while recombinant EXOG was diluted to 60 ng/µL in 10 mM sodium acetate buffer pH 5.5 and immobilised to a level of approximately 1000 RU. Two flow cells without immobilised ligand were included for background subtraction. All binding experiments were carried out in HBS-EP buffer (10 mM HEPES pH 7.6, 150 mM NaCl, 0.1 mM EDTA, 0.05% (v/v) Tween-20) at a constant flow rate of 30 µL/min. RNase H1 and its domain variants (full-length, catalytic domain, hybrid-binding domain) were dialysed against HBS-EP buffer and injected as analytes in serial dilutions (31.25 nM to 2 µM). Each cycle consisted of a 150 s association phase followed by a 350 s dissociation phase. Sensor surfaces were regenerated using sequential injections of 50 mM NaOH and 1 M NaCl containing 10 mM glycine pH 2.0. Sensorgrams were analysed using Biacore T200 Evaluation Software. Kinetic rate constants (k_{on} and k_{off}) and equilibrium dissociation constants (K_d) were determined by global fitting to a 1:1

Langmuir binding model. Fit quality was assessed by inspection of residuals and χ^2 values. All experiments were independently repeated at least twice.

6.6. Glycerol-gradient ultracentrifugation

To isolate EXOG-RNase H1 complexes, a 2-20% glycerol gradient was prepared using HN2% (20 mM HEPES pH 7.5, 140 mM NaCl, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, 2% (v/v) glycerol) and HN20% (same buffer with 20% glycerol). For each tube, 2 ml of HN2% was layered over 2 ml of HN20% in 4 ml Beckman clear ultracentrifuge tubes. Gradients were formed using a Gradient Master 108 (BioComp Instruments) and pre-equilibrated at 4 °C before use. EXOG and RNase H1 were dialysed together for 2 hours at 4 °C against Gradient Reaction Buffer (20 mM HEPES pH 7.5, 140 mM NaCl, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol). The following samples were prepared in 110 µl: (1) EXOG-RNase H1 complex (4 µM EXOG and 8.8 µM RNase H1), (2) EXOG alone (4 µM), and (3) RNase H1 alone (8.8 µM). Prior to loading onto the gradients, samples were incubated for 5 minutes on ice and centrifuged at 10,000 × g for 5 minutes at 4 °C to remove potential aggregates. Samples were carefully layered on top of the preformed gradients and subjected to ultracentrifugation at ~$2.15 \times 10^5 \times g$ for 5.5 hours at 4 °C using an SW60 Ti rotor (Beckman Coulter). After centrifugation, gradients were manually fractionated in 150 µl steps from the top. Collected fractions were analysed by SDS-PAGE.

6.7. DNA and RNA substrate preparation

All oligonucleotides used in this study were purchased from Metabion International AG and are listed in Table 1 below. Annealing of the substrates was performed by mixing the respective oligonucleotides in 20 mM Tris-HCl pH 8.1, 0.1 mM EDTA buffer, placing the tubes in boiling water, and allowing them to cool to room temperature. A 1:1 molar ratio was used for standard substrates, while in the case of fluorescently labelled substrates, a 1.2:1 ratio of unlabelled to labelled strands was applied to ensure complete annealing.

Substrate number	Oligo Name	Sequence (5' → 3')	Modification	Length (nt)
S1	MRS#51	GCTTTTGCTCCCGGTCACGGTATGTTTTTTTTTTGCGTGCATTATGTCAGATCA	Biotin (5' & 3')	54
	MRS#71	CGAAAACGAGGGCCAGTGCCATAC	Fluorescein (5')	24
	MRS#60	CGCACGUAUAUCAGUCUAGU	None	20
S2	MRS#500	CGCACGTCAGTTGTATGGCACTGGCCCTCGTTTTCG	Biotin (5')	42
	MRS#201	CGAAAACGAGGGCCAGTGCCATAC	ddC (3')	24
	MRS#240	CUGACGUGCG	None	10
S3	MRS#309	CGCACGTCAGTTTGTATGGCAGTGGCCCTCGTTTTCG	Biotin (5')	41
	MRS#499	CGAAAACGAGGGCCAGTGCCATAC	ddC (3')	24
	MRS#240	CUGACGUGCG	None	10
S4	MRS#305	CGCACGTCAGTTTTTTTTGTATGGCAGTGGCCCTCGTTTTCG	Biotin (5')	40
	MRS#201	CGAAAACGAGGGCCAGTGCCATAC	ddC (3')	24
	MRS#240	CUGACGUGCG	None	10
S5	MRS#343	ACTAGACTGTATTACGTTATACTAGACTGTATTACGTGCGTTTTTTTTTTGTATGGCACTGGCCCTCGTTTTCG	Biotin (5' & 3')	74
	MRS#315	CGAAAACGAGGGCCAGTGCCATAC	Cy5 (5')	24
	MRS#341	cgc acg uaa uac agu cua gu ATA ACG TAA TAC AGT CTA GT	FLSN (3')	40
S6	MRS#343	ACTAGACTGTATTACGTTATACTAGACTGTATTACGTGCGTTTTTTTTTTGTATGGCACTGGCCCTCGTTTTCG	Biotin (5' & 3')	74
	MRS#353	CGAAAACGAGGGCCAGTGCCATAC	Cy5 (5') & ddC (3')	24
	MRS#341	cgc acg uaa uac agu cua gu ATA ACG TAA TAC AGT CTA GT	FLSN (3')	40
S7	MRS#88	CGCACGTCAGTTTTTTTTGTATGGCACTGGCCCTCGTTTTCG	None	42
	MRS#201	CGAAAACGAGGGCCAGTGCCATAC	ddC (3')	24
	MRS#240	CUGACGUGCG	None	10
S8	MRS#41	ACTAGACTGTATTACGTGCGGTATGGCACTGGCCCTCGTTTTCG	Biotin (5' & 3')	44
	#129	CGAAAACGAGGGCCAGTGCCATAC	Biotin (5')	24
	MRS#63	cgcacguaauACAGTCTAGT	FLSN (3')	20
S9	MRS#415	CGCACGTCA	none	9
	MRS#414	ugacgugcg	none	9
S10	MRS#284	CGAAAACGAGGGCCAGTGCCATAC	Biotin (5') & ddC (3')	24
	MRS#50	ACTAGACTGTATTACGTGCGTTTTTGTATGGCACTGGCCCTCGTTTTCG	Biotin (5' & 3')	50
	MRS#64	cgc acg uaa uAC AGT CTA GT	none	20

Table 1. Nucleic acid substrates used in this study. Compilation of the oligonucleotide substrates employed across the experiments. The table lists the substrate identifier, the oligo name, the sequence written from 5′ to 3′, the presence of any modifications as reported in the table, and the strand length in nucleotides. Each grouped set of rows corresponds to one substrate composed of the listed strands.

6.8. Two-nuclease cleavage assay

Cleavage reactions were carried out in a total volume of 20 µL in HK140 buffer (20 mM HEPES pH 7.5, 140 mM KCl, 5% (v/v) glycerol, 1 mM TCEP), supplemented with 10 mM $MgCl_2$. Reactions contained 200 nM fluorescently labelled RNA/DNA hybrid substrate (Table S1) and recombinant human RNase H1 (wild-type or pathogenic variants), with or without recombinant EXOG. In the first experiment, RNase H1 was titrated at final concentrations of 25, 50, 100, and 200 nM, in the absence or presence of 100 nM EXOG. In the second experiment, EXOG was titrated at the same concentrations (25, 50, 100, and 200 nM) in the presence or absence of a constant 100 nM RNase H1. Each titration series was performed separately for wild-type RNase H1 and two pathogenic variants: RNase H1-V142I and RNase H1-A185V. Substrates were fluorescently labelled at the 3′ end of the downstream oligonucleotide with fluorescein. Reactions were incubated at 25 °C for 5 minutes and stopped by the addition of 60 µL STOP buffer (90% formamide, 50 mM EDTA pH 8.0), followed by heat denaturation at 95 °C for 4 minutes. Reaction products were resolved by denaturing polyacrylamide gel electrophoresis (urea-PAGE) using 20% acrylamide gels containing 7 M urea in TBE buffer (0.5×: 44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA). Gels were run at 45 W for 50 minutes. Fluorescent signals were detected using a Typhoon fluorescence imager (Cytiva). All reactions were performed in triplicate in three independent experiments.

6.9. Crystallisation trials

6.9.1. Trial 1: EXOG-H140G-CTD complex

The EXOG-CTD complex was assembled by mixing EXOG-H140G (stock concentration: 98 µM, 7.55 mg/ml) and CTD (273 µM, 4.37 mg/ml) to final concentrations of 50 µM (~3.85 mg/ml) and 55 µM (~0.88 mg/ml), respectively, corresponding to a 1:1.1 molar ratio. Proteins were combined in a buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, m0.5 mM EDTA and were supplemented with 10 mM $MgCl_2$. The total protein concentration in the final crystallisation mixture was approximately 4.7 mg/ml. Prior to complex assembly, both proteins were clarified by centrifugation (10 min, maximum speed, 4 °C). The complex was incubated on ice for 15 minutes before crystallisation setup. Crystallisation trials were performed using sitting-drop vapour diffusion in 96-well Intelli-Plates, set up at room temperature using a Mosquito® liquid-handling robot (TTP Labtech). Drops were dispensed in two protein-to-reservoir ratios: 1:1 (400 nL + 400 nL) and 2:1 (800 nL + 400 nL). The following seven commercial sparse-matrix screens were used: MORPHEUS I, II, III; JCSG Plus; HELIX (Molecular dimensions), and INDEX and Crystal

Screen (Hampton Research). Plates were incubated at 18 °C and inspected regularly by light microscopy.

6.9.2. Trial 2: EXOG-H140G-CTD complex

The EXOG-CTD complex was assembled by mixing EXOG-H140G and CTD to final concentrations of 50 µM (~3.85 mg/ml) and 110 µM (~1.76 mg/ml), respectively, corresponding to a 1:2.2 molar ratio. Proteins were mixed in HK140 buffer (20 mM HEPES pH 7.0, 140 mM KCl, 5% (v/v) glycerol, 1 mM TCEP) and supplemented with 10 mM $MgCl_2$. Two independent preparations were assembled for this condition. In the first preparation, EXOG-H140G (stock: 85 µM, ~6.55 mg/ml) and CTD (309 µM, ~4.94 mg/ml) were combined in a final volume of 580 µl. In the second, EXOG-H140G (92.5 µM, ~7.12 mg/ml) and CTD (303 µM, ~4.85 mg/ml) were mixed in 420 µl. The final protein concentration in each crystallisation mixture was approximately 5.6 mg/ml. Prior to mixing, each protein was clarified by centrifugation (10 min, maximum speed, 4 °C). The assembled complexes were incubated on ice for 30 minutes before crystallisation setup. Crystallisation trials were performed using sitting-drop vapour diffusion in 96-well Intelli-Plates, set up using a Mosquito® liquid-handling robot. Drops were dispensed in two protein-to-reservoir ratios: 1:1 (400 nL + 400 nL) and 2:1 (800 nL + 400 nL). Two sets of commercial screens were used in parallel trials: (1) JCSG Plus, HELIX, INDEX, and Crystal Screen, and (2) MORPHEUS I, II, and III. Plates were incubated at 18 °C and inspected regularly by light microscopy.

6.9.3. Trial 3: Optimisation of the EXOG-H140G-CTD complex

To refine conditions from initial sparse-matrix screening, a targeted optimisation trial was performed using the EXOG-CTD complex at a 1:2.2 molar ratio. Proteins were mixed to final concentrations of 50 µM (~3.85 mg/ml) EXOG-H140G and 110 µM (~1.76 mg/ml) RNase H1-CTD in HK140 buffer (20 mM HEPES pH 7.0, 140 mM KCl, 5% (v/v) glycerol, 1 mM TCEP), supplemented with 10 mM $MgCl_2$. The complex was clarified by centrifugation (10 min, maximum speed, 4 °C) and incubated on ice for 15 minutes prior to crystallisation setup. Optimisation was performed manually using the sitting-drop vapour diffusion method in 24-well crystallisation plates (Hamilton), with two full plates set up in parallel. Drops were dispensed at two protein-to-reservoir ratios: 1 µl + 1 µl and 2 µl + 1 µl. A custom matrix was prepared based on condition HELIX 39, using 0.05 M Bis-Tris pH 7.0 as buffer. Lithium sulfate was varied across rows at four concentrations (0.05 M, 0.1 M, 0.2 M, 0.3 M), and MPD was varied across columns at six concentrations (15%, 20%, 25%, 30%, 35%, 40% v/v). All stock solutions were freshly prepared; MPD stock was obtained from Molecular Dimensions. Plates were incubated at 18 °C and monitored regularly by light microscopy.

6.9.4. Trial 4: EXOG-H140G-RNase H1 complex

The EXOG-RNase H1 complex was assembled at a 1:2.2 molar ratio by mixing EXOG-H140G and RNase H1 to final concentrations of 40 µM (~3.08 mg/ml) and 88 µM (~2.55 mg/ml), respectively, in HN140 buffer (20 mM HEPES pH 7.0, 140 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP), supplemented with 10 mM $MgCl_2$. Two independent preparations were performed, in total volumes of 650 µl and 300 µl. The estimated total protein concentration in the crystallisation mixtures was approximately 5.6 mg/ml. Prior to mixing, both proteins were clarified by centrifugation (10 min, maximum speed, 4 °C), and the assembled complexes were incubated on ice for 15 minutes. Crystallisation trials were carried out using sitting-drop vapour diffusion in 96-well Intelli-Plates, set up with a Mosquito® liquid-handling robot. Drops were dispensed at two protein-to-reservoir ratios: 1:1 (200 nL + 200 nL) and 2:1 (400 nL + 200 nL). In the first round, a panel of ten commercial sparse-matrix screens was tested. In the second round, six additional screens were screened: Shotgun, PEGRx 1, and PEGRx 2 (Hampton Research), and PACT Premier, Wizard 1, and Wizard 2 (Molecular Dimensions). Plates were incubated at 18 °C and monitored regularly by light microscopy.

6.9.5. Trial 5: RNase H1-RNA/DNA substrate complex

The RNase H1-substrate complex was prepared by mixing wild-type full-length RNase H1 (stock concentration: 205 µM, ~5.95 mg/ml) and an RNA/DNA hybrid duplex substrate (S9, Tsble1) (500 µM, ~2.78 mg/ml) to final concentrations of 100 µM (~2.90 mg/ml) RNase H1 and 110 µM (~0.61 mg/ml) substrate, corresponding to a 1:1.1 molar ratio. The complex was assembled in a buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, 0.5 mM EDTA, and supplemented with 10 mM $MgCl_2$. Before assembly, RNase H1 was clarified by centrifugation (10 min, maximum speed, 4 °C). The assembled complex was incubated on ice for 15 minutes. Crystallisation trials were performed using sitting-drop vapour diffusion format using a Mosquito® robot and 96-well Intelli-Plates. Drops were dispensed at two protein-to-reservoir ratios: 1:1 (200 nL + 200 nL) and 2:1 (400 nL + 200 nL). Twelve commercial sparse-matrix screens were tested in total: Index, Natrix, Crystal Screen, Crystal Screen Lite, ShotGun Screen, and SaltRx (Hampton Research), and PEGRx 1 and 2, Wizard 1 and 2, HELIX, and JCSG-plus (Molecular Dimensions). Plates were incubated at 18 °C and inspected regularly by light microscopy.

6.10. Structure predictions

To investigate the structural basis of the interaction between human RNase H1 and EXOG, I used AlphaFold2-Multimer via the localColabFold platform (YoshitakaMo/localcolabfold) (Evans et al., 2022; Jumper et al., 2021; Mirdita et al., 2022). The modelling pipeline was run

on the Athena high-performance computing cluster provided by PLGrid (Cyfronet), which offers GPU-based computational nodes optimised for large-scale machine learning workloads. Predictions were carried out locally using the precompiled multimer model weights distributed with ColabFold. The modelling process employed the full suite of internal AlphaFold modules, including multiple sequence alignment generation via MMseqs2 in its default configuration, enabling fast and sensitive detection of homologous sequences. Template-based structure search was explicitly disabled, following the default settings of ColabFold, to prioritise de novo structural inference from sequence information alone. Sequence alignments were generated against the UniRef30 database distributed with the local ColabFold package, with no manually curated templates or external restraints applied. All assemblies were modelled as heteromeric complexes, an EXOG biological dimer and a monomeric RNase H1 or its fragment, in a single prediction run, without post hoc docking.

Four biologically relevant assemblies, each comprising the EXOG homodimer in complex with distinct RNase H1 fragments, were investigated:

- EXOG dimer + full-length RNase H
- EXOG dimer + CTD-HBD fragment
- EXOG dimer + CTD
- EXOG dimer + CTD-Linker

These constructs were chosen to reflect the modular architecture of RNase H1 and to assess which regions contribute most to direct interaction with EXOG. Two complementary prediction strategies were applied to each assembly to balance model diversity with prediction depth:

- Multi-seed model inference - To explore conformational variability and assess the stability of predicted protein-protein interfaces, 20 models were generated per complex using the --random-seed 42 --num-seeds 4 option, initiating four independent prediction trajectories with different random seeds, each producing five models. These random variations enabled sampling of a broad range of plausible interaction geometries and assessment of contacts recovered consistently across runs. All predicted structures were relaxed using the built-in AMBER force field to improve local stereochemistry by removing steric clashes and strained geometries. Following visual inspection and ranking based on combined ipTM and pTM scores, the highest-ranked model from the multi-seed set was retained for further evaluation.
- Extended recycling protocol - As a complementary approach, five additional models were generated per complex using a high-depth prediction strategy with 48 recycling

steps (--num-recycle 48). This allowed AlphaFold2 to iteratively refine the predicted structure over many passes through the network, often yielding higher-confidence predictions and more stable interfacial contacts. All five models were retained for further evaluation. AMBER-based relaxation was again applied to ensure stereochemical quality.

Model quality was assessed using AlphaFold internal confidence metrics: the predicted TM-score (pTM), reflecting the accuracy of overall model folding, and the inter-chain predicted TM-score (ipTM), reflecting the accuracy of predicted inter-protein orientation. Additionally, Predicted Aligned Error (PAE) plots were inspected to evaluate the positional uncertainty between chains and to identify regions of potential flexibility or poorly defined interfacial geometry. For each complex, six models in total, one from the multi-seed set and five from the extended recycling set, were subjected to interface analysis using PDBePISA (Krissinel & Henrick, 2007) and residue contact mapping. Interfacial contacts were defined as any pair of heavy atoms from different chains within 4.0 Å. Contact frequency profiles were compiled across the six models, and residues showing high-frequency interactions were designated as candidate determinants of the EXOG-RNase H1 interface for further analysis. Predicted structures were visualised in UCSF ChimeraX (v1.8).

6.11. Polymerase gap-filling assay

Polymerase gap-filling reactions were carried out in a total volume of 20 µL in HK140 buffer (20 mM HEPES pH 7.5, 140 mM KCl, 5% (v/v) glycerol, 1 mM TCEP) at 25 °C. Each reaction contained 50 nM fluorescently labelled DNA/RNA hybrid substrate (S1, Table 1), 1.5 µM RNase H1-D210N, 12.5 nM Pol γA, 25 nM Pol γB, 1 µM deoxynucleotide triphosphates (dNTPs), and 10 mM $MgCl_2$. The hybrid substrate contained a defined single-stranded gap flanked by a 5′ RNA primer and a 3′ downstream DNA oligonucleotide annealed to a complementary DNA template strand. The upstream oligonucleotide was labelled at the 5′ end with fluorescein to enable fluorescent detection. Reactions were initiated by the addition of dNTPs and incubated at 25 °C. At defined time points (0.25, 0.5, 1, 10, 15, and 30 minutes), reactions were stopped by adding 60µL STOP buffer (90% formamide, 50 mM EDTA pH 8.0), followed by a 4-minute incubation at 95 °C. Samples were separated by denaturing polyacrylamide gel electrophoresis (urea-PAGE) using 20% acrylamide gels containing 7 M urea in TBE buffer, run at 40 W for 38 minutes. Fluorescent signals were detected using a Typhoon fluorescence imager (Cytiva). All reactions were performed in triplicate and independently repeated three times.

6.12. Biolayer Interferometry

Biolayer interferometry (BLI) experiments were conducted to investigate the sequential binding of Pol γ and RNase H1 to a gapped DNA/RNA substrate (S2, S3, S4, Table 1). All bottom strands were 5′ biotinylated to enable their immobilisation on streptavidin-coated biosensors, and the top strand carried a 3′-terminal dideoxycytidine (ddC) modification to prevent strand extension by Pol γ and thereby trap the polymerase on the substrate. Streptavidin-coated SAX biosensors (Sartorius) were pre-hydrated in HK20 buffer (20 mM HEPES pH 7.5, 20 mM KCl, 1 mM TCEP) + 4% (v/v) glycerol, 400 µg/ml bovine serum albumin (BSA), 0.01% (v/v) Tween-20 for 10 minutes before use. All proteins and substrates were diluted in HK20 buffer to a final concentration of 1 µM in a total volume of 50 µL. The BLI experiments were performed on a BLItz system (fortebio) at room temperature with continuous mixing at 1500 rpm. The procedure consisted of the following steps: (1) equilibrating the biosensors in HK20 buffer for 30 s to establish baseline; (2) immobilising the biotinylated DNA/RNA substrate onto the biosensor for 500 s; (3) washing in HK20 buffer for 150 s to remove unbound substrate; (4) incubating the biosensor in a solution containing Pol γ (1 µM) for 550 s; (5) washing in HK20 buffer for 300 s to remove unbound Pol γ; (6) incubating in a solution containing RNase H1-D210N (1 µM) for 500 s; and (7) performing a final wash in HK20 buffer for 300 s to remove unbound RNase H1. The binding responses were recorded in real-time. Each experiment was performed in triplicate.

6.13. Coupled gap-filling-nuclease assay

Reactions were performed in a total volume of 20 µL in HK140 buffer (20 mM HEPES pH 7.5, 140 mM KCl, 5% (v/v) glycerol, 1 mM TCEP), supplemented with 10 mM $MgCl_2$ and 0.1 mg/mL bovine serum albumin (BSA). Where indicated, 100 µM deoxynucleotide triphosphates (dNTPs) were added to support gap-filling activity. Each reaction contained 200 nM DNA/RNA hybrid substrate (S5, S6 Table 1), increasing concentrations of RNase H1 variants (0.2, 2, 20, and 200 nM), 100 nM Pol γA, and 200 nM Pol γB. Substrates used in these assays were fluorescently labelled at the 3′ end of the downstream oligonucleotide with fluorescein. Reactions were incubated at 37 °C for 1 hour and stopped by the addition of 4× STOP buffer (90% formamide, 50 mM EDTA pH 8.0), followed by denaturation at 95 °C for 4 minutes. Samples were separated by denaturing polyacrylamide gel electrophoresis (urea-PAGE) using 20% acrylamide gels containing 7 M urea in TBE buffer, run at 40 W for 48 minutes. Fluorescent signals were visualised using a Typhoon fluorescence imager (Cytiva). Fluorescence signal intensities were analysed using ImageJ (https://imagej.net/ij/index.html) to quantify RNA cleavage. For each RNase H1 concentration, the intensity of the final cleavage product band was calculated as a percentage of the total lane intensity and corrected by

subtracting the corresponding value from the substrate-only control lane. This quantification was used to assess RNase H1 activity in the absence of Pol γ. In a separate analysis, reactions containing both RNase H1 and Pol γ were quantified using the same approach, and the corrected values were used to evaluate RNA cleavage in the presence of polymerase. In all cases, background cleavage observed in the absence of protein was subtracted to ensure accurate comparison across conditions. Data were obtained from three independent experiments. For each condition, quantification was performed in triplicate, and results were reported as mean ± standard deviation (SD).

6.14. Complex formation on a size-exclusion chromatography column

Size-exclusion chromatography (SEC) was used to analyse the formation of the Pol γ-RNase H1-RNA/DNA hybrid complex. The complex was assembled in a total volume of 200 µL by mixing preformed Pol γ holoenzyme, RNase H1, and the RNA/DNA hybrid substrate (S7, Table 1) at final concentrations of 2.5 µM, 7.5 µM, and 2.5 µM, respectively, corresponding to a 1:3:1 molar ratio, in SEC buffer (20 mM HEPES pH 7.5, 140 mM KCl, 5% (v/v) glycerol, 5 mM β-mercaptoethanol, 1 mM dATP, 5 mM $CaCl_2$). The reaction mixture was assembled by sequential addition of dATP, substrate, and $CaCl_2$, followed by the proteins. After incubation at 4 °C for 5 minutes, the sample was centrifuged at 12,000 × g for 10 minutes to remove aggregates. The clarified sample (200 µL) was injected onto a Superdex™ 200 10/300 GL column (GE Healthcare) pre-equilibrated in SEC buffer supplemented with 0.1 mM dATP. The chromatographic separation was performed on a Bio-Rad NGC system at a flow rate of 0.25 mL/min. Elution was monitored by UV absorbance at 280 nm, and fractions were collected for further analysis. Selected elution fractions were analysed by SDS-PAGE under denaturing conditions and stained with Coomassie Brilliant Blue to verify protein composition and assess complex formation.

6.15. Cryo-EM sample preparation and data acquisition

Complex 1

Complex 1 was prepared by glycerol gradient ultracentrifugation. The Reaction buffer contained 20 mM HEPES pH 7.5, 140 mM KCl, 10 mM $MgCl_2$, 1 mM dATP, and 5 mM β-mercaptoethanol. Gradients were formed in 4 ml Beckman clear tubes using a Gradient Master 108 (BioComp Instruments) from 2 ml HK2% buffer and 2 ml HK15% buffer, where HK2% and HK15% were the Reaction buffer supplemented with 2% or 15% glycerol, respectively. The gradients were equilibrated at 4 °C. The ternary complex contained gapped RNA/DNA hybrid substrate (S10, Table1) at a final concentration of 2 µM, Pol γ (exo-) at 2 µM, and RNase H1 at

4 μM. It was assembled by sequential addition of the substrate and Pol γ with a 5-minute incubation on ice, followed by the addition of RNase H1 and a further 5-minute incubation. The mixture was clarified at 14,000 × g for 5 minutes at 4 °C, diluted to 200 μl, and 100 μl were layered onto the prepared gradient. Ultracentrifugation was carried out in an SW60 Ti rotor at ~2.15×10^5 × g for 5 hours at 4 °C. Gradients were fractionated from the top in 150 μl steps, and fractions were analysed by SDS-PAGE. Fractions showing co-migration of Pol γ and RNase H1 were pooled. The pooled material was concentrated to approximately 50 μl using a 100 kDa molecular weight cutoff centrifugal device (Amicon, RC membrane). To remove glycerol, the sample was diluted with 600 μl of Reaction buffer and reconcentrated to about 40 μl. The final preparation was flash-frozen in liquid nitrogen and stored at -80 °C until thawed and applied to cryo-EM grids. Grids were Quantifoil R2/1 and were prepared in a Vitrobot system (Thermo Fisher Scientific) at 4 °C and 100% humidity with 3 μl sample applied, blot time 2 s, blot force 4, and wait time 2 s. Prepared grids were first screened on a Glacios microscope, and full data collection was performed on a Titan Krios microscope at the SOLARIS National Synchrotron Radiation Centre operating at 300 kV. Data were recorded using Falcon III and K3 direct electron detectors. In total 7,800 untilted micrographs and 2,926 micrographs at 30° stage tilt were acquired at a nominal magnification of 105,000×, corresponding to a calibrated pixel size of 0.84 Å. The defocus range was −1.5 to −0.6 μm. The total electron dose was 41.0 $e^-/Å^2$ fractionated over 40 frames. Data were processed in cryoSPARC software (Punjani et al., 2017; Rubinstein & Brubaker, 2015; Punjani et al., 2020; Punjani & Fleet, 2021; Punjani & Fleet, 2023).

Complex 2

Proteins were buffer exchanged into a buffer that contained 20 mM HEPES pH 7.5, 140 mM KCl, and 5 mM beta mercaptoethanol. Complexes were prepared on ice in a final volume of 20 μl to yield 2 μM Pol γ (exo-), 4 μM RNase H1, and 2 μM RNA/DNA substrate (S7, Table 1), supplemented with 5 mM $CaCl_2$ and 1 mM dATP. Order of addition was as follows: buffer, substrate, Pol γ, incubate on ice for 5 minutes, RNase H1, incubate on ice for 15 minutes. Mixtures were centrifuged at 14,000 g for 5 minutes at 4 °C, aliquoted, and stored at -80 °C until use. The oligonucleotide composition of the substrate is provided in Table X. Grids were prepared on Quantifoil R2/1 using a Vitrobot system from Thermo Fisher Scientific at 4 °C and 100% humidity with 3 μl sample applied, blot time 2 seconds, blot force 4, and wait time 2 seconds. Data were collected on a Glacios microscope at the SOLARIS National Synchrotron Radiation Centre operating at 200 kV with a spherical aberration of 2.7 mm. Movies were recorded in EER format at a pixel size of 0.73 Å per pixel with 623 frames per movie and a total exposure of 40.22 electrons per $Å^2$. Defocus values were set to −2.1, −1.8, −1.5, −1.2, and −0.9

μm. The dataset was collected at zero-degree stage tilt, and in total 2,041 micrographs were acquired. Data were processed in cryoSPARC software.

6.16. Molecular graphics and analysis in ChimeraX

Throughout this work, UCSF ChimeraX was used for molecular graphics and analysis, including inspection of cryo-EM maps, rigid body placement, distance and contact measurements, interface visualisation, and preparation of publication-quality figures. For predicted models, including those generated with AlphaFold, ChimeraX tools were applied to assess fit to density, local geometry, and residue-level contacts in a consistent manner across datasets (Pettersen et al., 2021, Goddard et al., 2018, and Meng et al., 2023).

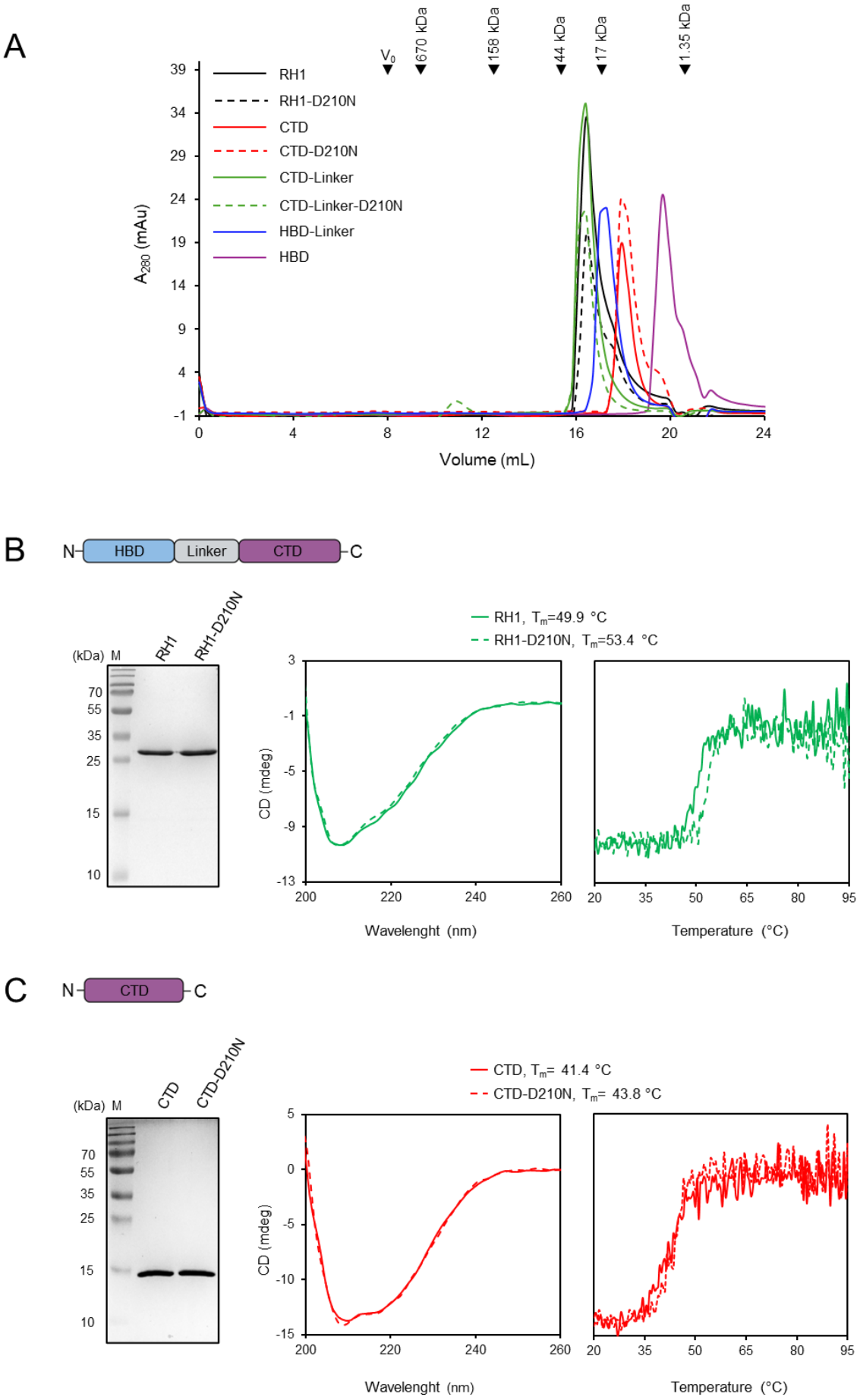
7. Results

7.1. Part I: Functional cooperation of RNase H1 and EXOG in mitochondrial RNA primer removal

Human mitochondrial DNA replication requires the complete removal of RNA primers to create ligatable DNA ends, yet the division of labor and coordination among the resident nucleases remain unclear. A reconstituted system showed that EXOG is indispensable for processing RNA containing intermediates and that its cleavage precedes DNA Ligase III activity, while RNase H1 provides complementary activity on RNA/DNA hybrids that alone does not complete primer removal (Karlowicz et al., 2022). However, that study did not establish whether RNase H1 and EXOG engage in direct physical interaction on shared substrates or how their activities are choreographed in space and time, a question that is particularly relevant given *in vivo* evidence that RNase H1 is essential for primer processing during initiation and completion of mitochondrial DNA replication. Therefore, I set out to define and test the direct coupling between RNase H1 and EXOG by integrating interaction assays, functional reconstitution, and structural analysis, with the goal of revealing the interfaces and mechanisms that enable efficient primer clearance and of assessing how disease-linked variants perturb this cooperation.

7.1.1. RNase H1 and EXOG purification and structural integrity assessment

To enable structure-function analyses of RNase H1 in the context of mitochondrial RNA primer removal, the full-length protein and its domain-specific variants were successfully overexpressed and purified. These included the N-terminal hybrid-binding domain (HBD), the C-terminal catalytic domain (CTD), and the corresponding linker-containing constructs. Analytical size-exclusion chromatography revealed that all RNase H1 variants were monomeric under the tested conditions, exhibiting distinct elution profiles in accordance with their molecular weights (Figure 11A-E). To assess the structural integrity and folding properties of all RNase H1 proteins, circular dichroism (CD) spectroscopy was performed (Figure 11B-E). The spectra confirmed well-defined secondary structures characteristic of α-helical content, consistent with known RNase H1 architectures. Thermal unfolding experiments revealed distinct melting temperatures (T_m) across the variants, reflecting differences in domain stability and folding.


A
V0
670 kDa
158 kDa
44 kDa
17 kDa
1.35 kDa
RH1
RH1-D210N
CTD
CTD-D210N
CTD-Linker
CTD-Linker-D210N
HBD-Linker
HBD
A280 (mAu)
Volume (mL)
B
N
HBD
Linker
CTD
C
(kDa) M
RH1
RH1-D210N
RH1, Tm=49.9 °C
RH1-D210N, Tm=53.4 °C
CD (mdeg)
Wavelenght (nm)
Temperature (°C)
C
N
CTD
C
(kDa) M
CTD
CTD-D210N
CTD, Tm= 41.4 °C
CTD-D210N, Tm= 43.8 °C
CD (mdeg)
Wavelenght (nm)
Temperature (°C)

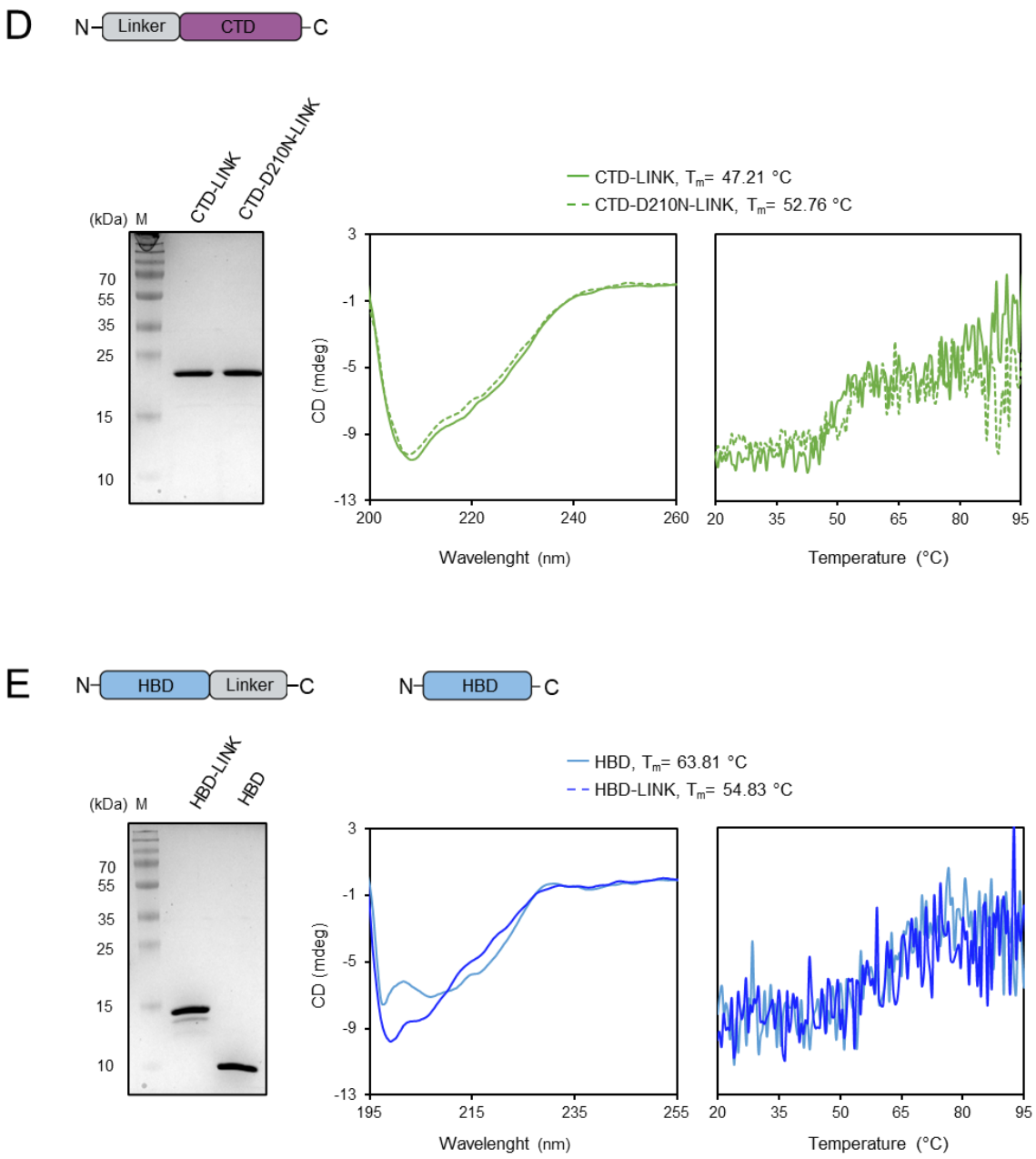


Figure 11. Purification and characterisation of RNase H1 and its variants. (A) Size exclusion chromatogram of RNase H1 variants separated on a Superdex 200 Increase 10/300 GL SEC column: wild-type RNase H1 (RH1) (black solid), RH1-D210N (black dashed), RNase H1 catalytic domain (CTD) (red solid), CTD-D210N (red dashed), CTD-Linker (green solid), CTD-D210N-Linker (green dashed), HBD-Linker (blue solid) and RNase H1 hybrid-binding domain (HBD) (purple solid). (B, C, D, E) SDS-PAGE analysis of purified RNase H1 variants (1 µg per lane) stained with Coomassie Brilliant Blue. Panels B, C and D use 15% gels to resolve full-length and CTD variants, respectively, while panel E shows HBD-Linker and HBD resolved on a 16% gel. Molecular weight marker (M) is indicated. (CD analysis) Circular dichroism spectra and thermal unfolding curves were recorded for RNase H1 variants (2.5 µM) in potassium phosphate buffer (20 mM KH_2PO_4/K_2HPO_4, 140 mM KCl, pH 7.4). Ellipticity at 208 nm was monitored from 20 °C to 95 °C to determine melting temperatures (T_m), which are indicated above the respective plots.

Additionally, thermal unfolding experiments revealed substantial differences in stability across the variants. The hybrid-binding domain (HBD) showed the highest melting temperature (T_m = 63.8 °C), indicative of its compact and thermodynamically stable structure. When linked to the flexible linker region, the stability of HBD decreased (T_m = 54.8 °C), suggesting partial destabilisation upon extension (Figure 11E). In contrast, the CTD displayed lower stability (T_m = 41.4 °C), which increased modestly when fused with the linker (T_m = 47.2 °C) or upon introduction of the D210N mutation (T_m = 43.8 °C and 52.7 °C for CTD and CTD-Linker, respectively) (Figure 11C-D). Interestingly, the catalytically inactive full-length RNase H1 (D210N) was markedly more stable (T_m = 53.4 °C) than the wild-type protein (T_m = 49.9 °C), indicating that the mutation might increase conformational rigidity (Figure 11B). These results reflect domain-specific contributions to folding cooperativity and thermal resilience. Together, these biophysical analyses confirm that all RNase H1 constructs are properly folded and biochemically stable, supporting their suitability for downstream interaction studies and functional assays. To investigate the structural consequences of disease-associated RNase H1 mutations, I characterised two pathogenic variants, V142I and A185V, previously linked to mitochondrial encephalomyopathy and impaired mtDNA replication (Al-Behadili et al., 2018). Both variants were successfully overexpressed and purified, although the overall yield and purity were notably lower than for the wild-type enzyme, indicating reduced solubility or conformational stability. Analytical size-exclusion chromatography confirmed that both variants were monomeric in solution and displayed sharp, symmetrical elution profiles similar to wild-type RNase H1 (Figure 12A), indicating proper folding and the absence of aggregation. Circular dichroism (CD) spectra confirmed the preservation of α-helical secondary structure, comparable to wild-type RNase H1 (Figure 12B), suggesting that global folding was retained. However, thermal denaturation experiments revealed a substantial loss of structural stability. The A185V mutant exhibited a markedly reduced melting temperature (T_m = 32.1 °C), while the V142I variant did not yield a discernible unfolding transition, indicating severe conformational destabilisation under native conditions (Figure 12B). These findings are in line with previous reports showing that both mutations, located near the catalytic site, disrupt hydrophobic packing critical for the structural integrity of the active site region and RNA/DNA hybrid binding (Al-Behadili et al., 2018). Together, these results suggest that the pathogenicity of these RNase H1 variants likely arises from compromised structural stability, which may impair their enzymatic function and hinder interactions with other mitochondrial replication components.

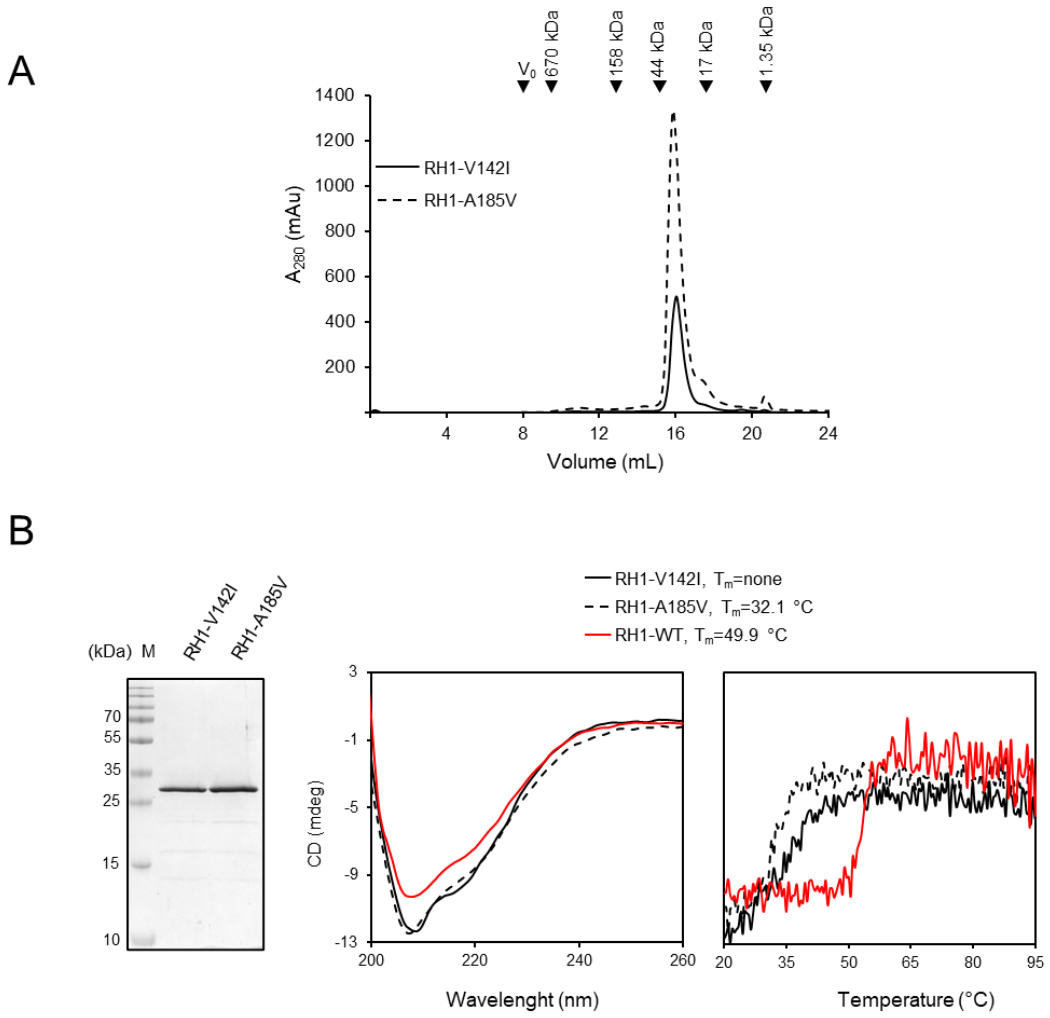


Figure 12. Purification and characterisation of RNase H1 pathogenic variants. (A) Size exclusion chromatogram of RNase H1 pathogenic variants separated on a Superdex 200 Increase 10/300 GL SEC column : RH1-V142I (black solid), RH1-A185V (black dashed). (B) SDS-PAGE (15%) analysis of purified RNase H1 pathogenic variants (1 µg per lane) stained with Coomassie Brilliant Blue. Molecular weight marker (M) is indicated. (CD analysis) Circular dichroism spectra and thermal unfolding curves were recorded for both proteins (2.5 µM) in potassium phosphate buffer (20 mM KH_2PO_4/K_2HPO_4, 140 mM KCl, pH 7.4). Ellipticity at 208 nm was monitored from 20 °C to 95 °C to determine melting temperatures (T_m), which are indicated above the plots.

EXOG is a mitochondrial 5′-exonuclease that contributes to mtDNA maintenance by processing DNA intermediates during human mitochondrial base excision repair (BER) and RNA primer removal (Wu et al., 2019; Karlowicz et al., 2022; Szymanski et al., 2017). Although its enzymatic activity is well established, the structural basis for its function and interactions with other components of the mitochondrial replication machinery remains less understood. To evaluate the structural integrity and folding of EXOG proteins used in this study, both the wild-type truncated variant (EXOG Δ58) and its catalytically inactive mutant (EXOG-H140G) were overexpressed

and purified for biophysical characterisation (Figure 13). Analytical size-exclusion chromatography revealed symmetrical elution profiles for both constructs, consistent with a homogeneous oligomeric state under native conditions (Figure 13A). Far-UV circular dichroism spectra confirmed well-defined secondary structure content for both variants, indicative of proper folding (Figure 13B). Thermal unfolding profiles showed cooperative transitions with melting temperatures of 42.8 °C for the wild-type and 40.9 °C for the H140G mutant, respectively (Figure 13B). These data demonstrate that the H140G mutation does not significantly destabilise the overall EXOG fold and confirm that both variants are suitable for downstream interaction and functional assays.

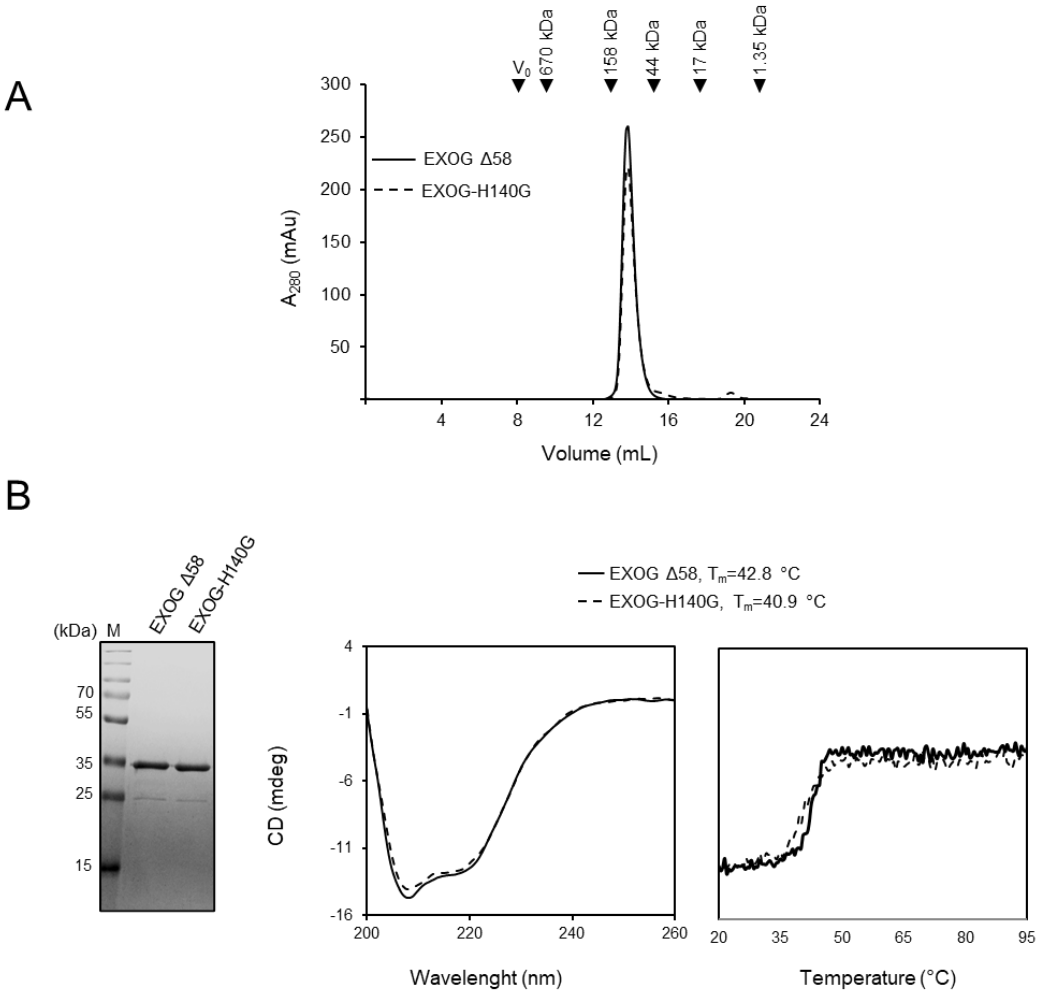
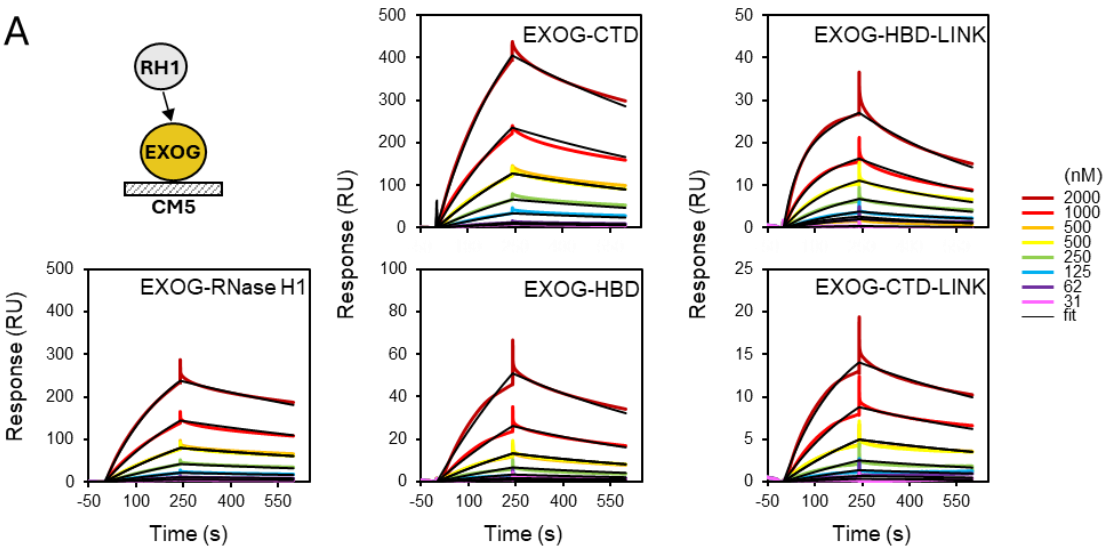


Figure 13. Purification and characterisation of EXOG Δ58 and EXOG-H140G. (A) Size exclusion chromatogram of EXOG Δ58 and EXOG-H140G separated on a Superdex 200 Increase 10/300 GL SEC column: EXOG Δ58 (black solid), EXOG-H140G (black dashed). **(B)** SDS-PAGE (15%) analysis of purified EXOG Δ58 and EXOG-H140G (1 µg per lane) stained with Coomassie Brilliant Blue. Molecular weight marker (M) is indicated. (CD analysis) Circular dichroism spectra and thermal unfolding curves were recorded for both proteins (2.5 µM) in potassium phosphate buffer (20 mM KH_2PO_4/K_2HPO_4, 140 mM KCl, pH 7.4). Ellipticity at 208 nm was monitored from 20 °C to 95 °C to determine melting temperatures (T_m), which are indicated above the plots.

7.1.2. Domain mapping of RNase H1-EXOG interaction

To investigate which regions of RNase H1 contribute to its binding to EXOG, I conducted surface plasmon resonance (SPR) experiments using a CM5 sensor chip. Figure 14 shows the resulting sensograms and kinetic parameters for five RNase H1 variants injected over immobilised EXOG. EXOG was covalently immobilised on the CM5 surface via standard amine coupling. Increasing concentrations of RNase H1, CTD, HBD, as well as CTD-Linker and HBD-Linker variants were injected as analytes. Each variant was tested in three independent titrations. Binding curves were fitted globally using a 1:1 Langmuir model to extract kinetic parameters.

A



B

Analyte	k_{on} (10^3 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	K_d (10^{-6} M)
RNase H1	2.48 ± 1.09	0.81 ± 0.09	0.362 ± 0.128
CTD	1.44 ± 0.63	0.97 ± 0.08	0.77 ± 0.31
CDT-LINK	2.3 ± 1.15	0.97 ± 0.39	0.5 ± 0.27
HBD-LINK	5.6 ± 5.6	2.09 ± 0.65	1.425 ± 1.906
HBD	0.27 ± 0.33	1.353 ± 0.125	12.6 ± 10.1

Figure 14. Surface plasmon resonance reveals domain-dependent interactions between RNase H1 and EXOG. (A) SPR sensorgrams showing the binding of RNase H1 variants to immobilised EXOG. Increasing concentrations of full-length RNase H1, CTD, CTD-Linker, HBD-Linker and HBD (31.25, 62.5, 125, 250, 500, 1000, and 2000 nM) were injected over the sensor chip surface, and binding responses were recorded. (B) Kinetic parameters derived from the interactions shown in (A). Association rate constants (k_{on}), dissociation rate constants (k_{off}), and equilibrium dissociation constants (K_d) were determined using a 1:1 binding model. Values represent the mean ± standard deviation from at least two independent experiments with global fitting.

Full-length RNase H1 displayed the strongest interaction with EXOG, characterised by a dissociation constant (K_d) of 0.38 ± 0.13 µM, an association rate constant (k_{on}) of 2.48 ± 1.09 × 10^3 M^{-1} s^{-1}, and a dissociation rate constant (k_{off}) of 0.81 ± 0.09 × 10^{-3} s^{-1}. Truncation into individual domains markedly influenced binding behaviour. The CTD alone retained measurable affinity (K_d = 0.77 ± 0.31 µM; k_{on} = 1.44 ± 0.83 × 103 M^{-1} s^{-1}; k_{off} = 0.97 ± 0.08 × 10^{-3} s^{-1}), suggesting that this domain may account for a substantial part of the recognition interface. Addition of the native linker to CTD (CTD-Linker) was associated with a modest improvement in affinity (K_d = 0.50 ± 0.27 µM) and a higher k_{on} (2.30 ± 1.15 × 10^3 M^{-1} s^{-1}), pointing to a potential role of the linker in facilitating complex formation. In contrast, the HBD domain alone

exhibited weaker binding (K_d = 12.6 ± 10.1 µM) with a low k_{on} (0.27 ± 0.33 × 10^3 M^{-1} s^{-1}) and a relatively fast k_{off} (1.35 ± 0.13 × 10^{-3} s^{-1}). However, fusion of HBD with the native linker (HBD-Linker) improved affinity by nearly an order of magnitude (K_d = 1.43 ± 1.91 µM), largely due to a pronounced increase in k_{on} (5.61 ± 5.59 × 10^3 M^{-1} s^{-1}), although this was accompanied by a higher k_{off} (2.09 ± 0.65 × 10^{-3} s^{-1}). Overall, these results reveal a consistent trend in which the catalytic domain of RNase H1 contributes most substantially to EXOG binding. The presence of the native linker appears to enhance binding in both domains, primarily by increasing the association rate, which may reflect a structural or conformational role in facilitating recognition. While the full-length protein achieves the highest affinity, the observed patterns across variants suggest that interdomain cooperation and linker-mediated flexibility could be important factors in achieving optimal interaction.

7.1.3. Glycerol gradient co-sedimentation suggests EXOG-RNase H1 complex formation

To investigate whether EXOG and RNase H1 can form a stable complex in solution, I performed glycerol gradient ultracentrifugation followed by SDS-PAGE analysis of the collected fractions (Figure 15A-C). The aim of this experiment was to determine whether the two proteins co-sediment in the gradient, which would be consistent with complex formation. Each panel of the figure includes both the SDS-PAGE analysis and a schematic representation of the expected behaviour during centrifugation. The following samples were prepared: EXOG alone (4 µM), RNase H1 alone (8.8 µM), and the EXOG-RNase H1 complex (4 µM EXOG and 8.8 µM RNase H1). For each condition, thirteen fractions were collected sequentially from the top (fraction 1) to the bottom (fraction 13) of the gradient.

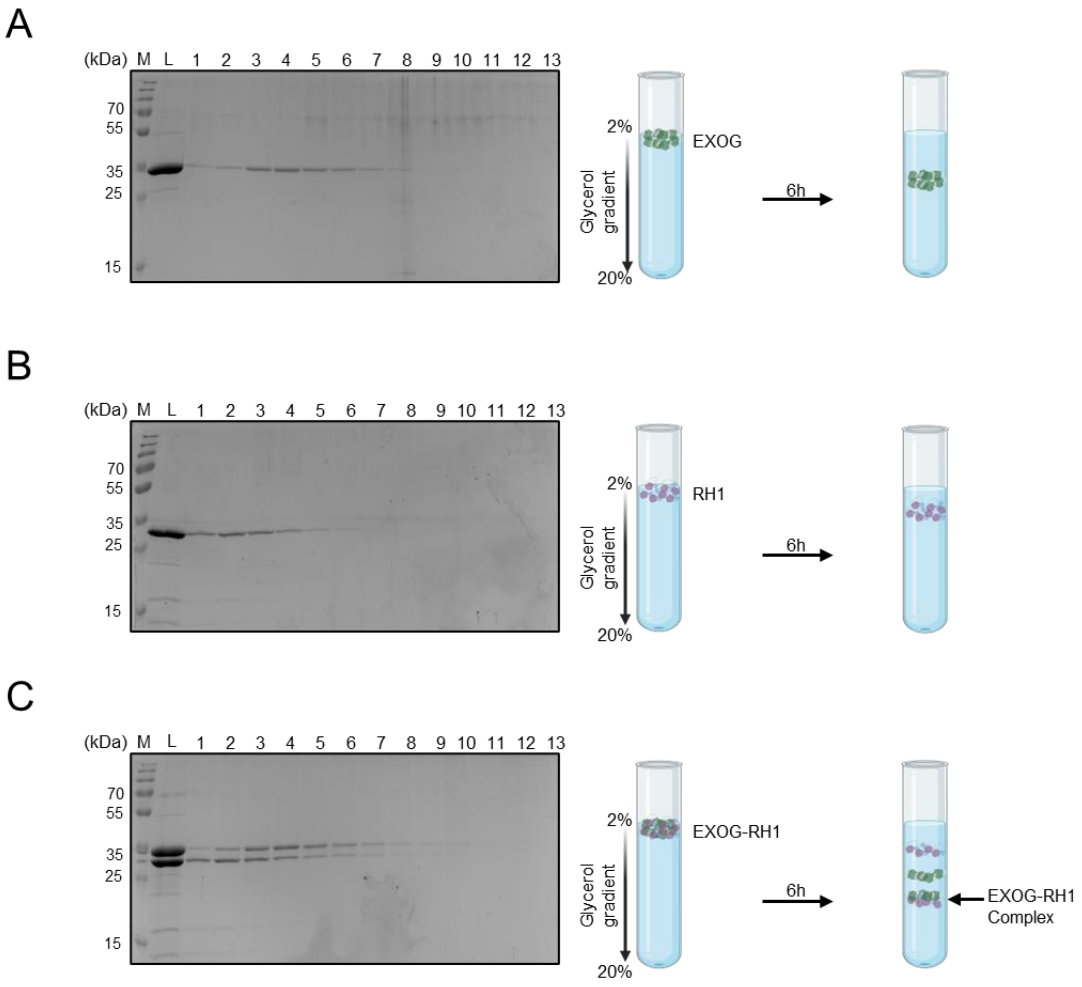


Figure 15. Glycerol gradient ultracentrifugation reveals co-sedimentation of EXOG and RNase H1. (A-C) 2-20% glycerol gradient ultracentrifugation followed by SDS-PAGE was used to assess whether EXOG and RNase H1 form a stable complex in solution. Individual proteins were analysed separately (A: EXOG alone; B: RNase H1 alone) and together (C: EXOG-RNse H1 mixed prior to loading). For each condition, thirteen fractions were collected from top to bottom and analysed by SDS-PAGE. Representative SDS-PAGE results are shown alongside schematic illustrations of protein sedimentation profiles across the collected fractions.

As shown in Figure 15A, EXOG alone sedimented predominantly in fractions 3 to 5, whereas RNase H1 alone (Figure 15B) displayed a band distribution spanning fractions 1 to 4. When both proteins were incubated and loaded onto the same gradient (Figure 15C), RNase H1 extended beyond its usual range and was detected in fractions 5 to 7, whereas EXOG largely retained its position with a low intensity tail into heavier fractions. In the mixed sample, the strongest overlap of both proteins occurred in fractions 5 to 7. The RNase H1 bands appeared less intense than those of EXOG, consistent with EXOG being a dimer that migrates as a monomer in SDS-PAGE and therefore contributing a higher molar amount per complex, while RNase H1 is monomeric. The redistribution of RNase H1 together with the overlap of both sedimentation profiles supports

the formation of an EXOG-RNase H1 complex. The limited resolving power of the gradient and the small mass difference between the free proteins and the potential complex likely reduce the magnitude of the shift. These findings are consistent with the previously established interaction between EXOG and RNase H1. Although glycerol gradient ultracentrifugation cannot determine the precise stoichiometry or binding mode, the observed redistribution provides evidence that a complex can form under the tested conditions. Based on these results, the method was initially considered a promising strategy for promoting efficient complex formation and for potential preparative use.

7.1.4. Two-nuclease system efficiently removes RNA primers

To functionally examine the cooperation between RNase H1 and EXOG in RNA primer processing, I performed a two-nuclease cleavage assay using a 10RNA/10DNA chimeric substrate with a single nick (S8, Table 1), as previously established by Karłowicz et al. (2022). This minimal substrate reproduces a simplified replication intermediate in which RNase H1 initiates cleavage of the RNA moiety, leaving behind 1-3 ribonucleotides which must be removed to complete primer processing. My original aim was to extend this system to more complex substrates mimicking mitochondrial replication intermediates, such as circular ssDNA templates primed with short RNA stretches. However, these attempts proved technically challenging and yielded inconsistent results. Therefore, I adopted the validated linear substrate model to systematically assess the functional interplay between RNase H1 and EXOG. As a first step, I reconstituted the two-nuclease cleavage system using wild-type RNase H1. Although the biochemical cooperation between RNase H1 and EXOG has been previously reported, my goal was to reproduce these findings under my experimental conditions and establish a reference point for subsequent analysis of disease-associated RNase H1 variants. The two-nuclease cleavage assay designed to assess the cooperation between RNase H1 and EXOG in removing RNA primers from a nicked RNA/DNA hybrid substrate is shown (Figure 16A,B). Titrations with wild-type RNase H1 and EXOG are shown (Figure 16C,D).

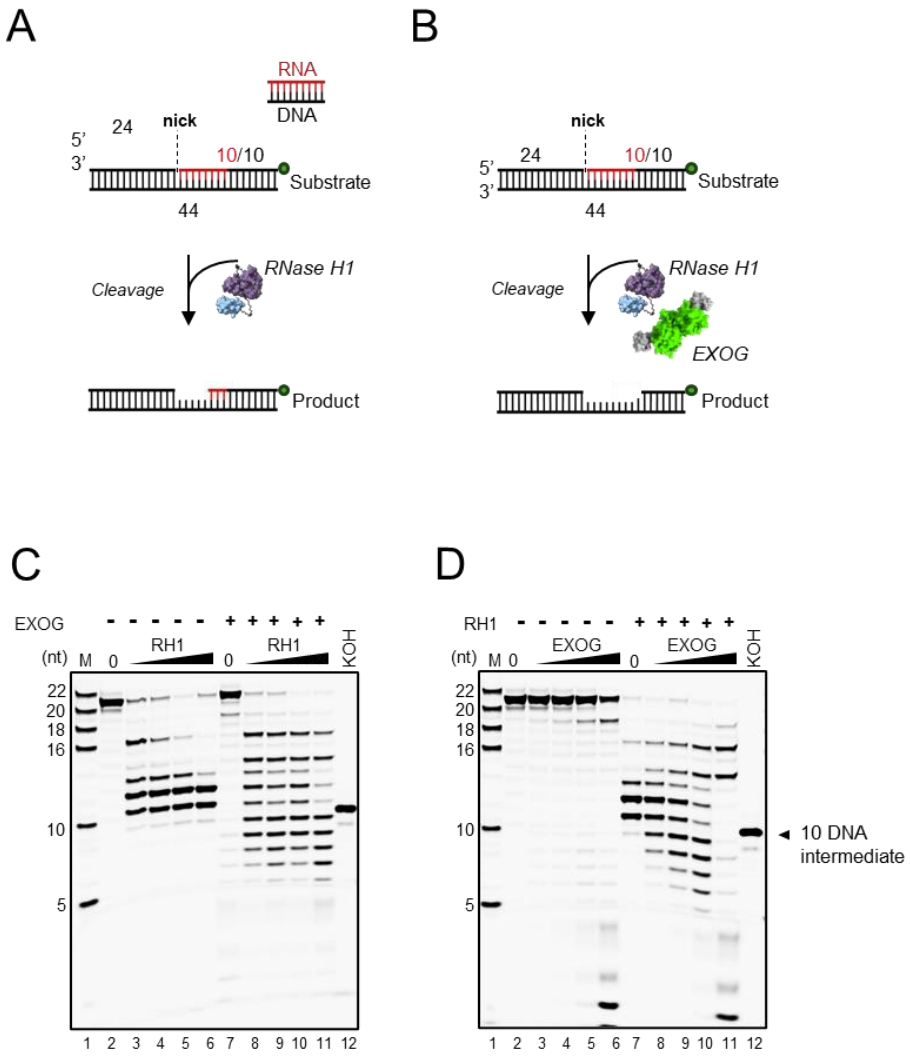


Figure 16. **Cleavage of a nicked RNA/DNA substrate by RNase H1 and EXOG.** (A, B) Schematic representations of RNA primer removal by RNase H1 alone (A) or in combination with EXOG (B). (C, D) Denaturing-PAGE analysis of cleavage of a fluorescently labelled nicked RNA/DNA substrate (200 nM). Panel C shows a titration of wild-type RNase H1 (25-200 nM) in the absence or presence of 100 nM EXOG, whereas panel D shows a titration of EXOG (25-200 nM) in the absence or presence of 100 nM RNase H1. Reactions were performed for 5 min at 25 °C in HK140 buffer supplemented with 10 mM $MgCl_2$ and stopped by addition of formamide-containing stop buffer. Reaction products were resolved by 20% urea-PAGE and visualised by fluorescence imaging. KOH indicates a chemical RNA hydrolysis control (300 mM KOH, 55 °C, 120 min). M, molecular weight markers.

Increasing concentrations of RNase H1 led to progressive degradation of the RNA moiety, generating a distinct banding pattern consistent with partial cleavage leaving 1-3 ribonucleotides at the RNA-DNA junction (Figure 16C, lanes 3-6). When 100 nM EXOG was added to these reactions, this residual RNA was efficiently removed, as indicated by the appearance of shorter cleavage products (Figure 16C, lanes 8-11). These results indicate that EXOG processes the

RNA fragments left behind by RNase H1, enabling complete primer clearance. The sequential nature of this cooperation is evident, because EXOG alone shows only marginal activity on the full-length substrate, particularly at lower concentrations, whereas the addition of constant RNase H1 (100 nM) significantly enhanced the cleavage efficiency across the EXOG titration (Figure 16D, lanes 3-6 and 8-11).

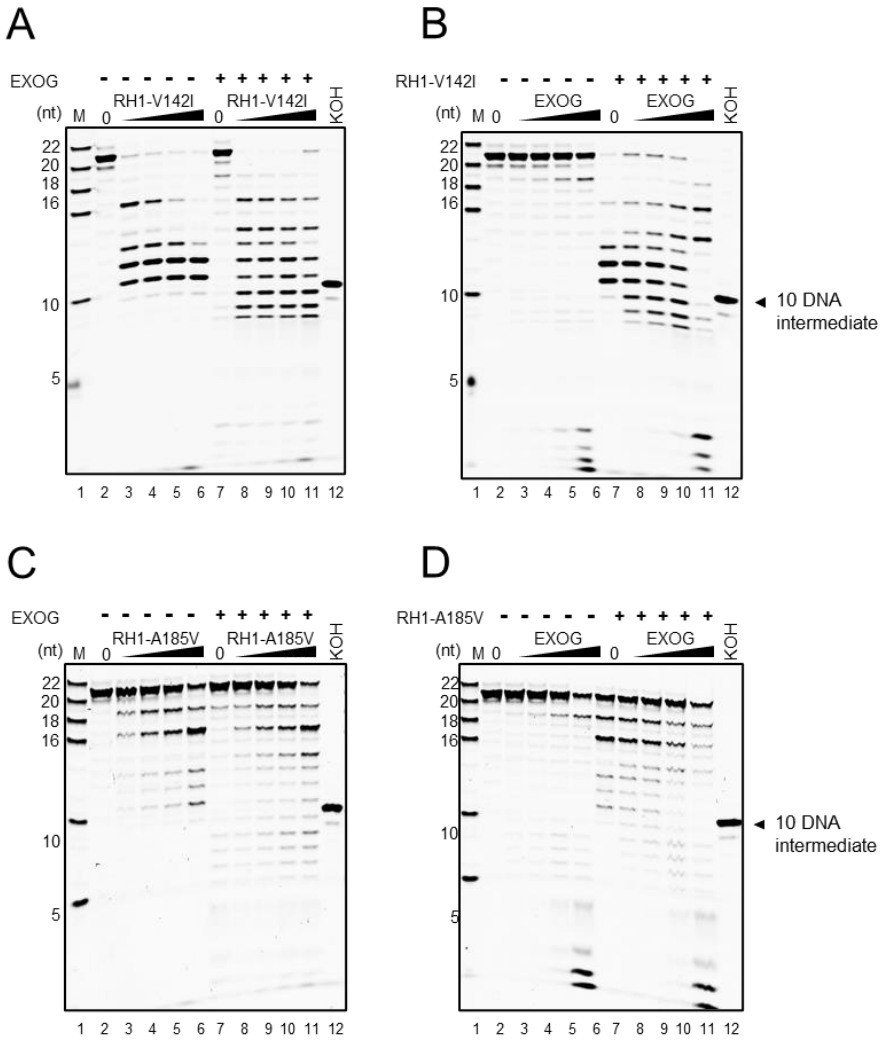


Figure 17. **Cleavage of a nicked RNA/DNA substrate by disease causing RNase H1 and EXOG**. (A, B) Denaturing-PAGE analysis of a fluorescently labelled nicked RNA/DNA substrate (200 nM) subjected to *in vitro* cleavage by variant RNase H1 proteins. Panels show titrations of the disease-associated mutant RNase H1-V142I (A) and RNase H1-A185V (B) across a 25-200 nM concentration range, performed either alone or in the presence of 100 nM EXOG. Reactions were incubated for 5 min at 25 °C in HK140 buffer supplemented with 10 mM $MgCl_2$ and terminated by addition of formamide-containing stop buffer. Reaction products were resolved on 20% urea-PAGE and visualised by fluorescence imaging. KOH indicates a control reaction with chemical RNA hydrolysis (300 mM KOH, 55 °C, 120 min). M, molecular weight markers.

The same experimental setup was applied to the pathogenic RNase H1 variant V142I (Figure 17A,B). In contrast to prior reports suggesting compromised activity (Al Behadili et al., 2018), this variant displayed cleavage kinetics comparable to the wild-type enzyme. Titration of RNase H1 V142I yielded a banding pattern indistinguishable from the wild-type (Figure 17A, lanes 3-6), and the addition of EXOG resulted in complete removal of the RNA segment, with product profiles matching those observed for wild-type RNase H1 (Figure 17A, lanes 8-11). This functional preservation was further confirmed in the reciprocal EXOG titration (Figure 17B), where the presence of RNase H1 V142I again supported efficient RNA degradation. These findings indicate that, despite the reduced stability, the catalytic activity of RNase H1 V142I is retained under the conditions of this assay.

In contrast, the A185V variant displayed a marked impairment in RNA cleavage (Figure 17C,D). RNase H1 A185V exhibited minimal activity across the tested concentrations, with only faint cleavage products visible at the highest enzyme level (Figure 17C, lanes 3-6). However, when EXOG was added to the titration series, substrate degradation became more apparent, with the emergence of lower molecular weight products corresponding to complete RNA removal (Figure 17C, lanes 8-11). These products were less intense than those for wild-type or V142I, suggesting that EXOG can only partially rescue the defective processing activity of A185V. A similar trend was observed in the complementary titration (Figure 17D, lanes 3-6 and 8-11), where EXOG alone performed poorly, but its cleavage efficiency improved significantly upon the addition of RNase H1 A185V. Altogether, these results confirm that RNase H1 and EXOG act in a cooperative manner to remove RNA primers from replication like substrates. While the wild-type and V142I variant both generate intermediates that are efficiently resolved by EXOG, the A185V mutation results in substantially reduced RNase H1 activity. Nevertheless, EXOG retains the capacity to process partially cleaved substrates even in this deficient context, underscoring its compensatory role in primer processing.

7.1.5. Crystallisation of EXOG-RNase H1 complex

The combined evidence from the SPR experiments and glycerol gradient ultracentrifugation indicated that RNase H1 and EXOG can engage in a direct and stable interaction in solution. These findings provided a clear rationale to pursue crystallographic studies aimed at resolving the structural basis of this association. A high-resolution structure of the complex would offer mechanistic insight into how both nucleases coordinate their activities, and would complement the biochemical results obtained in the earlier sections. Initial crystallisation trials were carried out using the EXOG-H140G-CTD complex at a 1:1.1 molar ratio (Materials and Methods, Crystallisation: Trail 1). The complex, at a final concentration of approximately 4.7 mg/ml, was

subjected to extensive matrix screening using seven commercial screens: Morpheus I, II, III, JCSG Plus, HELIX, INDEX, and Crystal Screen. Drops were dispensed at 1:1 and 2:1 protein-to-reservoir ratios using a Mosquito® robot and incubated at 18 °C. This initial screen encompassed hundreds of crystallisation conditions and yielded numerous hits with microcrystals. However, the crystals were consistently too small or poorly formed for reproducible optimisation. Aside from these microcrystalline precipitates, no other crystal-like objects were observed across the screens. The high density of weak or non-optimisable hits, without any promising lead condition, prompted a modification of the complex stoichiometry. A second screening trial was therefore undertaken using an increased CTD ratio.

To improve upon the inconclusive outcome of the initial crystallisation trials, the EXOG-H140G-CTD complex was mixed at a 1:2.2 molar ratio, at a final concentration of approximately 5.6 mg/ml (Materials and Methods, Crystallisation: Trail 2). Two independent rounds of crystallisation screening were performed using a total of seven commercial screens: JCSG Plus, HELIX, INDEX, Crystal Screen, and Morpheus I-III. Drops were set up at 1:1 (400 nL + 400 nL) and 2:1 (800 nL + 400 nL) protein-to-reservoir ratios and incubated at 18 °C. These extended trials once again produced a high number of precipitate-rich conditions and microcrystals, often with salt-like crystals. For over a year, none of the observed crystal forms showed potential for optimisation or data collection. However, approximately 1.5 years after the initial screens, a few conditions yielded better-defined crystals. Eight of the most promising candidates were selected and sent for X-ray diffraction analysis at a synchrotron. Of these, only one was confirmed not to be a salt crystal, but it diffracted to worse than 11 Å resolution, rendering it unsuitable for further analysis. This modest signal nevertheless motivated a focused optimisation campaign centred around the condition that had produced the single diffracting crystal. A summary of the identified conditions and representative images of the crystals are provided below (Figure 18).

A

1-23 1-34 1-39

B

HELIX position	Main components (mother liquor)	Crystals selected	Diffraction outcome
1-23	0.1 M NaCl, 0.005 M $MgCl_2$, 0.05 M Bis-Tris pH 7.0, 25% MPD	2	No diffraction
1-34	0.1 M $(NH_4)_2SO_4$, 0.002 M spermine tetrahydrochloride, 0.05 M Bis-Tris pH 7.0, 35% MPD	1	No diffraction
1-39	0.2 M Li_2SO_4, 0.05 M Bis-Tris pH 7.0, 20% MPD	4	One crystal diffracted >11 Å

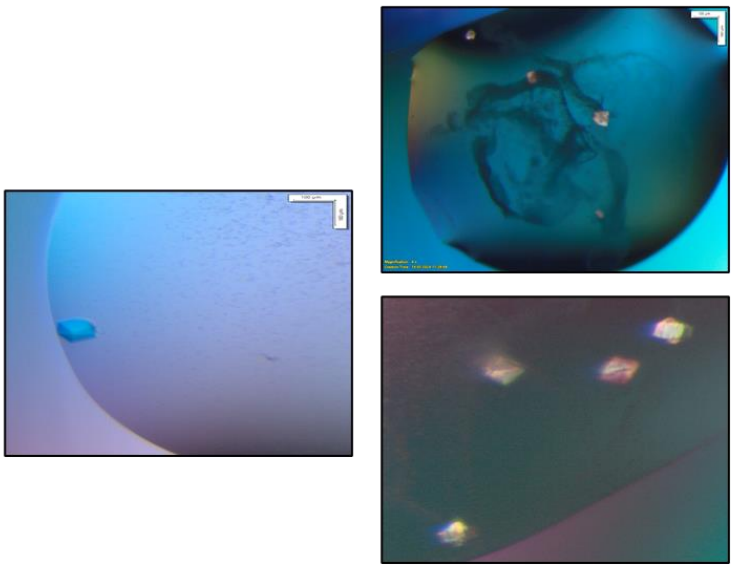
Figure 18. Representative crystals of the EXOG-CTD complex obtained from the HELIX screen. (A) Images show crystal formation observed in wells 1-23, 1-34, and 1-39 of the HELIX screen. Crystallisation drops were prepared using a 1:2.2 molar ratio of EXOG-H140G to CTD, at a final protein concentration of approximately 5.6 mg/ml. All shown crystals were obtained from drops set up at a 1:1 ratio of protein complex to reservoir solution (400 nL + 400 nL) and incubated at 18 °C. (B) Table below summarises the chemical composition of the corresponding mother liquors, the number of crystals selected for testing, and the diffraction outcomes for each condition.

Following the single diffracting crystal obtained under HELIX 39 conditions in Trial 2, a focused optimisation screen was performed. The original condition (0.2 M lithium sulfate, 0.05 M Bis-Tris pH 7.0, 20% MPD) was systematically modified by varying the concentrations of salt and precipitant across a defined matrix. Two 24-well plates were prepared manually using sitting-drop vapour diffusion at 18 °C, with drops assembled at 1:1 and 2:1 protein-to-reservoir ratios (Materials and Methods, Crystallisation: Trail 3). Despite systematic exploration of this condition space, no crystals were observed under any of the tested conditions.

To evaluate whether the full-length RNase H1 could form a crystallisable complex with EXOG-H140G, an extensive matrix screening was conducted at a 1:2.2 molar ratio (Materials and Methods, Crystallisation: Trail 4). The complex was prepared at a final concentration of approximately 6.0 mg/ml and screened against fourteen commercial crystallisation kits: Index, Natrix, PEGRx 1 and 2, Wizard 1 and 2, HELIX, Crystal Screen, Crystal Screen Lite, SaltRx,

ShotGun Screen, and JCSG-plus. Screens were set up using 96-well sitting-drop vapour diffusion at 18 °C, with drops dispensed at both 1:1 and 2:1 protein-to-reservoir ratios. Many conditions produced fine needle-like crystals, and two conditions yielded well-formed individual crystals suitable for X-ray diffraction. A total of two crystals were harvested and shipped to the synchrotron for data collection. The selected conditions are listed in the table in Figure 19.

A



Morpheus III 1-13

B

Screen&position	Main components (mother liquor)	Crystals selected	Diffraction outcome
Index 85	0.2 M $MgCl_2 \cdot 6H_2O$, 0.1 M Tris pH 8.5, 25% w/v PEG 3,350	1	No diffraction
Morpheus III 1-13	1.5% Vitamin mix; 3% w/v Sodium L-ascorbate, 3% w/v Choline chloride, 3% v/v D-Panthenol, 3% w/v Pyridoxine hydrochloride, 3% w/v Thiamine hydrochloride; 0.1 M Imidazole/MES pH 6.5; 40% v/v PEG 500 MME, 20% w/v PEG 20,000	1	2,3 Å

Figure 19. Crystallisation of the EXOG-RNase H1 complex and data collection summary. (A) Representative crystals obtained during Trial 4 from two conditions that yielded single crystals suitable for diffraction testing. Crystallisation was performed at a 1:2.2 molar ratio of EXOG-H140G to RNase H1, at a final protein concentration of approximately 6.0 mg/ml. All shown crystals were grown in sitting-drop vapour diffusion format at 18 °C using a 2:1 ratio of protein complex to reservoir solution (800 nL + 400 nL). (B) The table summarises the chemical composition of the mother liquors, number of harvested crystals, and diffraction results. One crystal, grown in Morpheus III condition 1-13, diffracted to 2.3 Å and was used for full data collection.

One of the crystals obtained in Trial 4 from the Morpheus III 1-13 condition diffracted to a resolution of 2.3 Å and was selected for full data collection. The resulting dataset was processed in space group P2(1) (4), and the corresponding data statistics are summarised in Table 2. Due to anisotropy and poor quality of the highest-resolution shell, the resolution cutoff was set to 2.6 Å. The dataset showed overall high completeness (98.2%) and redundancy (3.6), with a signal-to-noise ratio (I/σ(I)) of 8.77 and Rmeas of 9.0%. These values deteriorated substantially in the highest-resolution shell, with I/σ(I) dropping to 0.31 and Rmeas increasing to 398.1%. Analysis with Xtriage revealed translational non-crystallographic symmetry (tNCS), which guided the downstream structure solution strategy. Molecular replacement was carried out using Phaser as implemented in PHENIX with a range of search models, including EXOG monomer and dimer, full-length RNase H1, its catalytic domain, and various combinations thereof. All solutions converged on the EXOG dimer as the only identifiable entity in the asymmetric unit. No interpretable electron density attributable to RNase H1 was detected, most likely due to the high flexibility of the full-length protein arising from its linker region. The second crystal from Trial 4 did not produce any diffraction.

Space group	P2(1) (4)
Unit cell parameters [Å]	73.41 83.07 75.55 90.0 114.64 90.0
Resolution limit [Å]	40.21-2.43 (2.58-2.43)
No. of reflections	219077 (36085)
No. of uniques	60189 (9729)
Multiplicity	3.64 (3.71)
I/sigI	8.77 (0.31)
R_meas [%]	9.0 (398.1)
Completeness [%]	98.2 (98.7)
B(Wilson) [A^2]	84.82
Mosaicity [deg]	0.184
CC(1/2)	99.9 (35.9)
ISa	24.28

Table 2. Data collection and processing statistics for the EXOG-RNase H1 complex. Summary of diffraction data collected from a single crystal grown in Morpheus III condition 1-13 during Trial 4. Data were collected at a resolution range of 40.2-2.3 Å and processed in space group P2(1) with unit cell parameters as indicated. Values in parentheses correspond to the highest-resolution shell.

To obtain structural insight into full-length human RNase H1, a crystallisation trial was performed using its complex with a 9-mer RNA/DNA hybrid (S9, Table 1). The complex was prepared at final concentrations of 100 µM RNase H1 (~2.9 mg/ml) and 110 µM substrate (~0.6 mg/ml) (Materials and Methods, Crystallisation: Trail 5). A total of twelve commercial sparse-

matrix screens were tested using sitting-drop vapour diffusion at 18 °C, with drops set at both 1:1 and 2:1 protein-to-reservoir ratios. Despite the broad condition coverage and inclusion of the hybrid substrate to promote stability and order, no crystal formation was observed. One likely explanation is that the protein concentration was suboptimal, falling below the levels that had previously yielded crystal hits with EXOG-containing complexes.

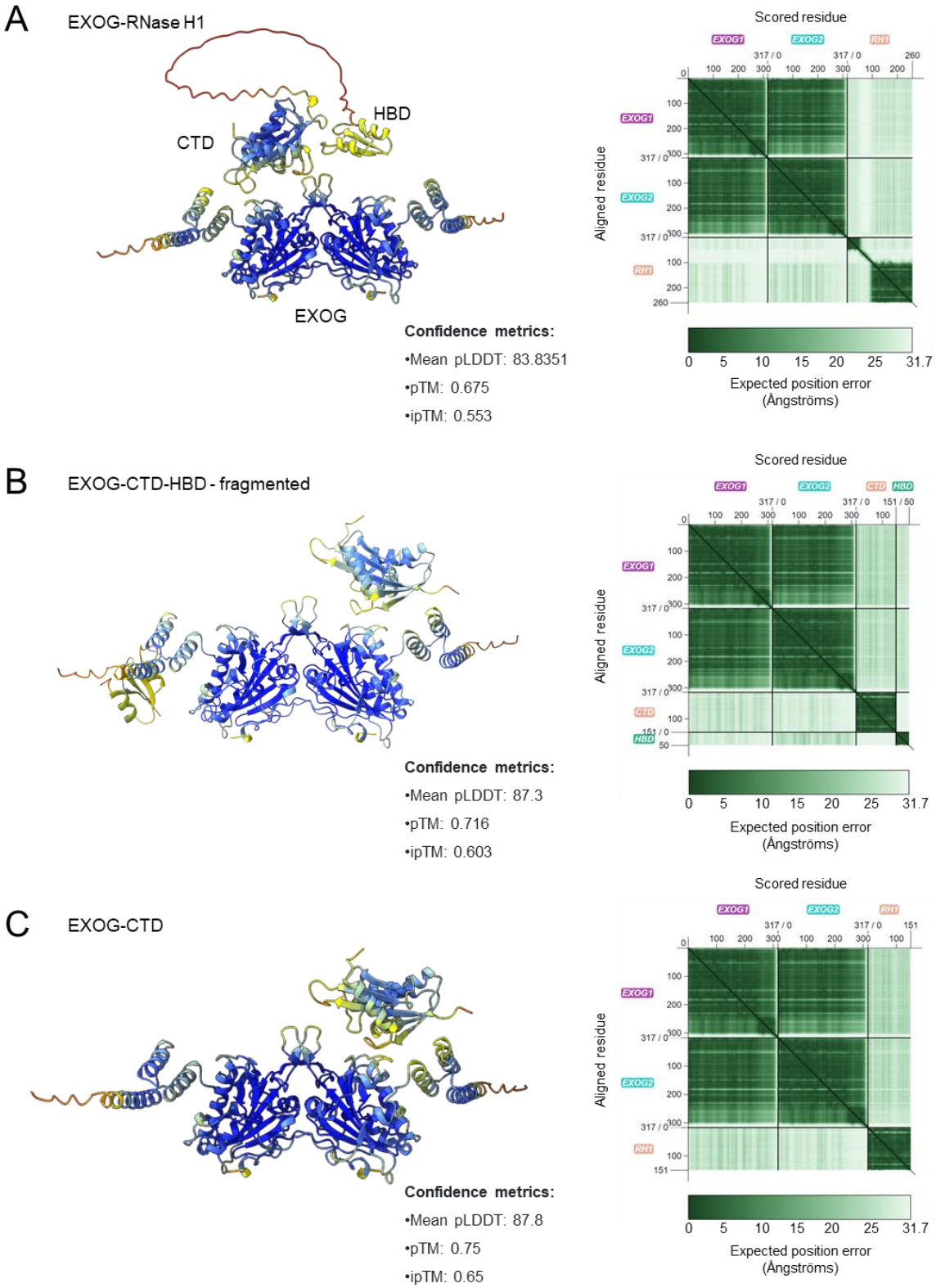
7.1.6. Structure prediction and identification of EXOG-RNase H1 interface residues

Given the unsuccessful attempts to crystallise the EXOG-RNase H1 complex, I employed computational structure prediction to identify candidate residues involved in the interaction. The rationale was to gain structural insight into potential binding interfaces, which could then be tested experimentally using site-directed mutagenesis and binding assays. Using AlphaFold2-Multimer implemented via localColabFold, I modelled four biologically relevant assemblies comprising the EXOG homodimer in complex with different RNase H1 fragments: (A) full-length RNase H1 (residues 27-286), (B) the CTD-HBD fragment, (C) the CTD alone, and (D) the CTD-linker construct. For all assemblies except CTD-HBD, two complementary prediction strategies were applied: (i) stochastic multi-seed sampling (20 models per complex), from which the best-ranked model was selected, and (ii) an extended recycling protocol (5 models per complex), all retained for downstream analysis. The CTD-HBD construct was predicted only using the extended recycling strategy, and solely the top-ranked model (rank_1) was included in further analysis due to substantial inconsistencies in domain orientation and interfacial contacts among the other models. Predictions involving EXOG in complex with the HBD domain alone were excluded from downstream analysis, as none of the generated models exhibited any inter-protein contacts.

All selected models displayed high confidence in both local and global prediction metrics (Figure 20). The full-length RNase H1 complex (Figure 20A) yielded a mean pLDDT of 83.8, with a pTM of 0.675 and ipTM of 0.553. The CTD-HBD model (Figure 20B) had a pLDDT of 87.3, pTM of 0.716, and ipTM of 0.603. The CTD-only complex (Figure 20C) scored highest, with pLDDT of 87.8, pTM of 0.750, and ipTM of 0.650. Finally, the CTD-linker construct (Figure 20D) produced a pLDDT of 83.8, a pTM of 0.709, and an ipTM of 0.594.

In AlphaFold-Multimer, pLDDT provides a per-residue estimate of local accuracy, while pTM reflects the global topology of the full model. Notably, ipTM is specifically designed to evaluate the predicted accuracy of inter-chain orientations, making it the most relevant confidence metric for interpreting predicted protein-protein interfaces. In addition, Predicted Aligned Error (PAE) plots (Figure 20, right panels) were examined for each assembly to assess positional uncertainty between the EXOG and RNase H1 chains. Low inter-chain PAE values in the CTD (Figure 20C) and CTD-linker (Figure 20D) complexes supported the high ipTM scores and indicated well-

defined docking geometries. In contrast, the full-length RNase H1 assembly (Figure 20A) exhibited elevated inter-chain PAE in the HBD and linker regions, suggesting substantial flexibility and reduced certainty in their relative positioning. The CTD-HBD model (Figure 20B) showed moderate inter-chain PAE, consistent with a stable CTD interface but variable HBD orientation.



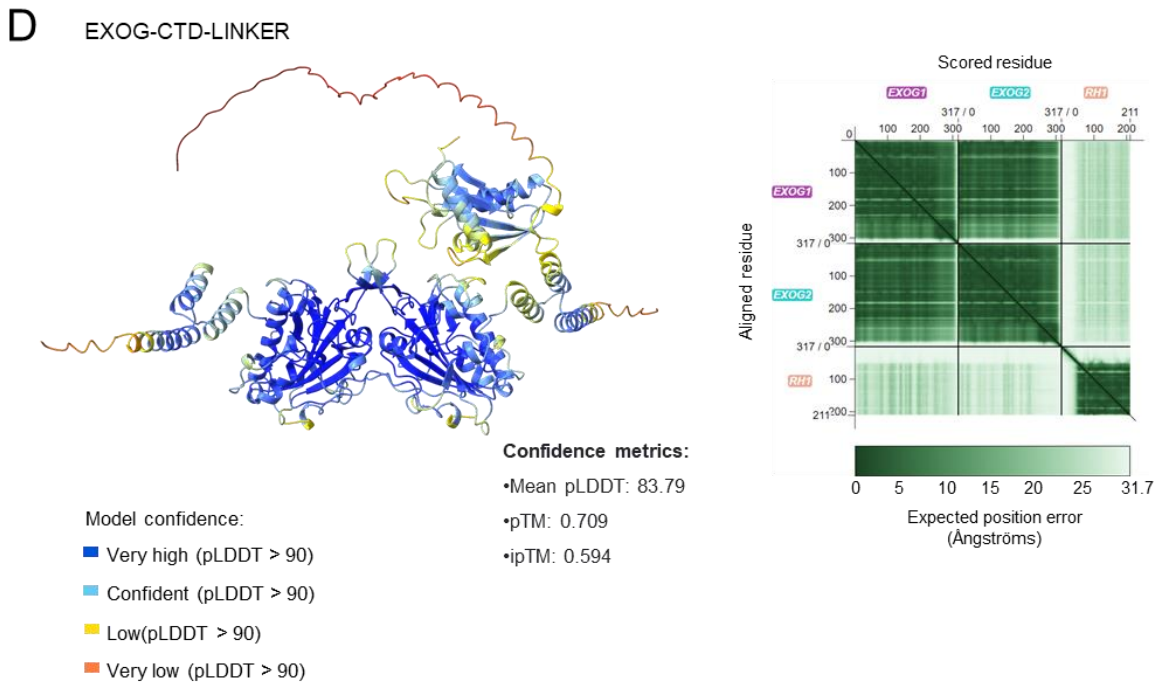


Figure 20. Structural models of the EXOG homodimer in complex with RNase H1 and its variants, predicted using AlphaFold2-Multimer. (A-D) show top-ranked (rank_1) models from extended-recycling predictions (5 models per complex) for four EXOG-RNase H1 assemblies: (A) full-length RNase H1, (B) CTD-HBD fragment, (C) CTD alone, and (D) CTD-linker construct. tructural models are coloured by per-residue pLDDT confidence and paired with corresponding predicted aligned error (PAE) plots. Each complex was modelled using AlphaFold2-Multimer via localColabFold, without templates. Confidence metrics (mean pLDDT, pTM, ipTM) are shown below each model.

The consistently high ipTM values (>0.55) observed across most assemblies, together with the PAE-derived confidence patterns, suggested the presence of potentially stable binding modes. These models warranted detailed structural analysis of the interfacial regions using contact mapping, PDBePISA metrics, and contact frequency profiling across models. Comparative analysis revealed that the CTD-only complex yielded the highest ipTM score (0.650), indicating the most confident prediction of interface geometry. This was followed by the CTD-HBD (ipTM = 0.603) and CTD-linker (ipTM = 0.594) constructs, both of which also demonstrated stable docking poses with well-defined contacts. In contrast, the full-length RNase H1 assembly showed the lowest interface confidence (ipTM = 0.553), despite having the largest molecular surface. These results suggest that the catalytic domain of RNase H1 likely plays a central role in EXOG binding, while additional regions such as the hybrid-binding domain or linker may contribute secondarily to modulating interface context or flexibility.

To characterise the predicted interaction surfaces between EXOG and RNase H1, I analysed inter-chain contacts across all six models per complex, except for the CTD-HBD construct, for which only the top-ranked model was included due to the inconsistencies described above. Contact analysis was conducted by manual inspection of each structure using UCSF ChimeraX, where interfacial residues were defined as those forming heavy-atom contacts within a 4.0 Å distance threshold. All models were additionally submitted to PDBePISA to identify hydrogen bonds and salt bridges between subunits, providing a chemically meaningful assessment of interface formation. Contacts lacking plausible spatial orientation or chemical feasibility were excluded. The remaining contacts were aggregated across models, and their frequency of occurrence was calculated to highlight consistently recurring interactions. The resulting contact frequency map (Figure 21) revealed several interaction hotspots. The most frequently involved residues from EXOG were D169, D207, G257, Q259, N331, N334, E336, while key RNase H1 contact residues included G137, F139, K204, K236, Q257, H260, S265. These regions are therefore considered strong candidates for constituting the core binding interface.

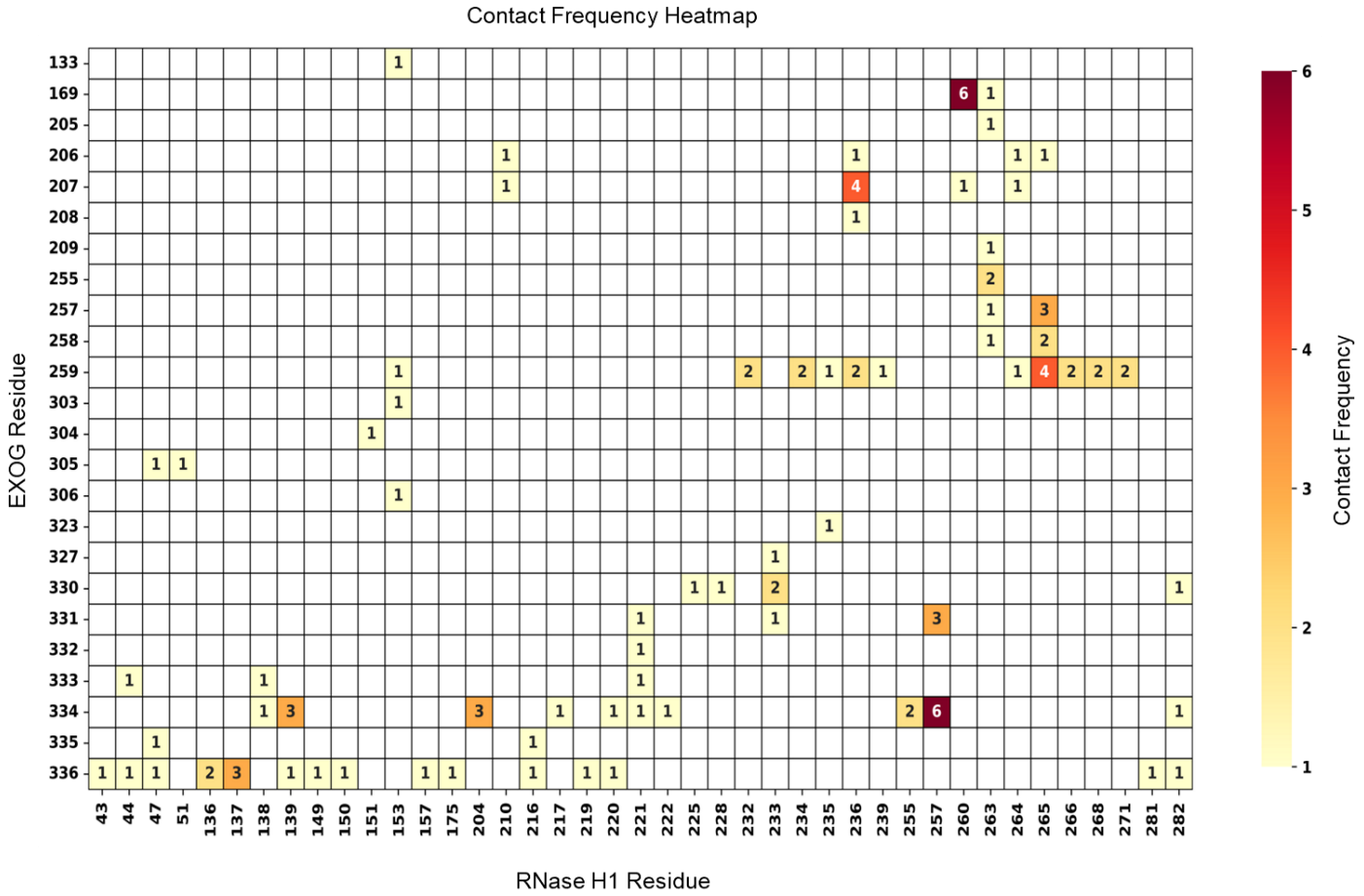


Figure 21. Residue-level contact frequency map of EXOG-RNase H1 predicted complexes. The heatmap summarises all non-redundant inter-chain contacts identified across six predicted EXOG-RNase H1 models (five from extended recycling, one from multi-seed sampling), combining structural inspection in UCSF ChimeraX and chemical annotation by PDBePISA. Contacts were curated by visual evaluation of AlphaFold2-Multimer models using a 3.0 Å cutoff for sidechain-heavy atom distances and were included only if spatially and chemically plausible. Each cell in the matrix represents a pair of residues forming a contact, with the number indicating how many times that contact was independently identified, either by AlphaFold-based inspection or PDBePISA, in any of the analysed models. The intensity of the colour corresponds to this cumulative frequency.

To further prioritise candidate interface residues for mutagenesis, I focused exclusively on the two highest-ranked models per complex, one from the multi-seed ensemble and one from the extended recycling prediction. A contact frequency map generated from these two models alone (Figure 22A) revealed four recurrent residue pairs: EXOG D169, G257, N331, N334 and RNase H1 Q257, H260, S265, with the most prominent interactions involving EXOG D169-RNase H1 H260 and EXOG N334-RNase H1 Q257. Restricting the analysis to inter-chain contacts classified as hydrogen bonds or salt bridges by PDBePISA in these models (Figure 22B) further narrowed the list to EXOG residues D169, N331, N334 and RNase H1 residues Q257, H260 as consistently involved in chemically plausible interactions. These residues were therefore selected as targets for site-directed mutagenesis and carried forward for functional validation.

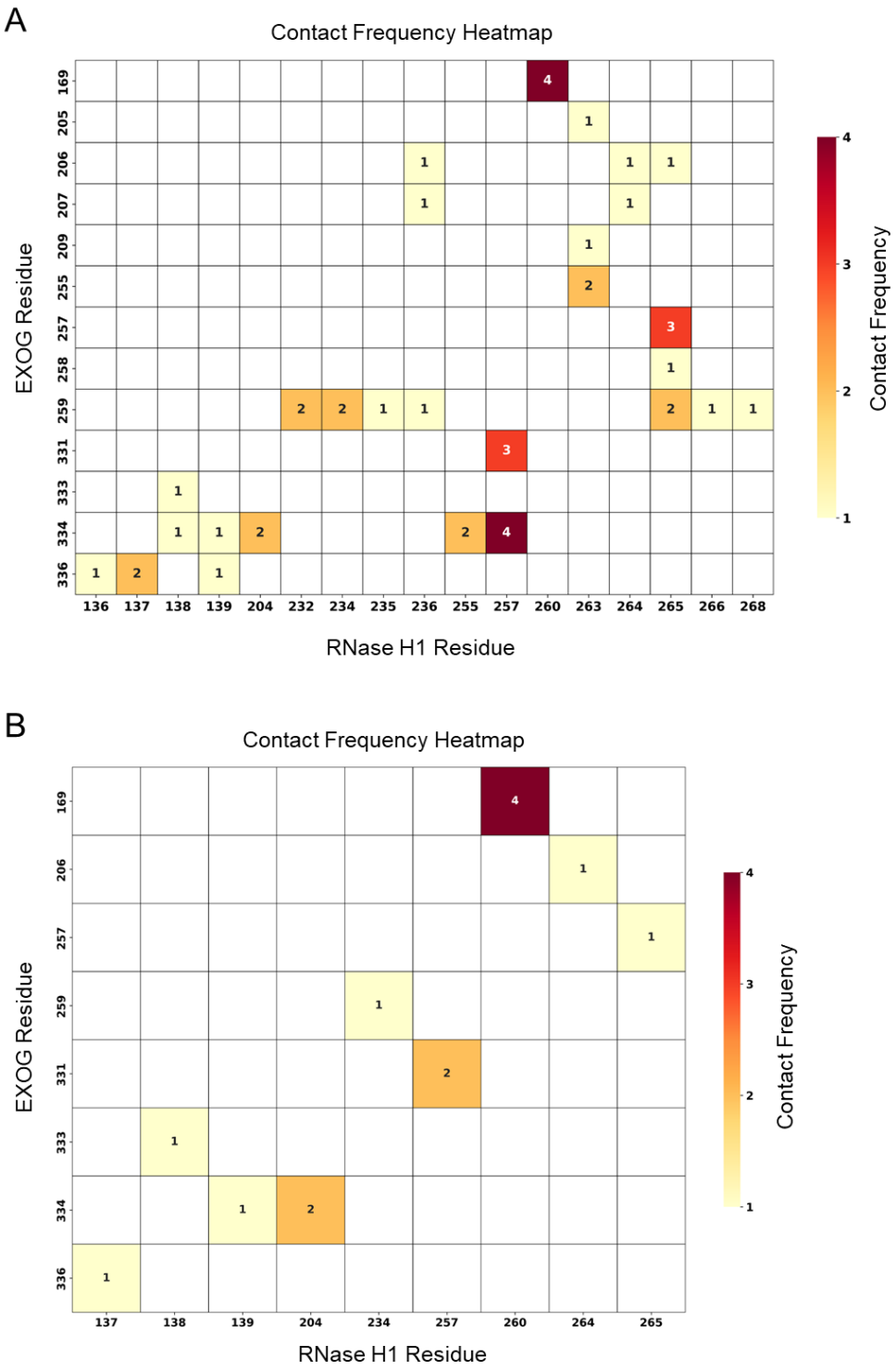
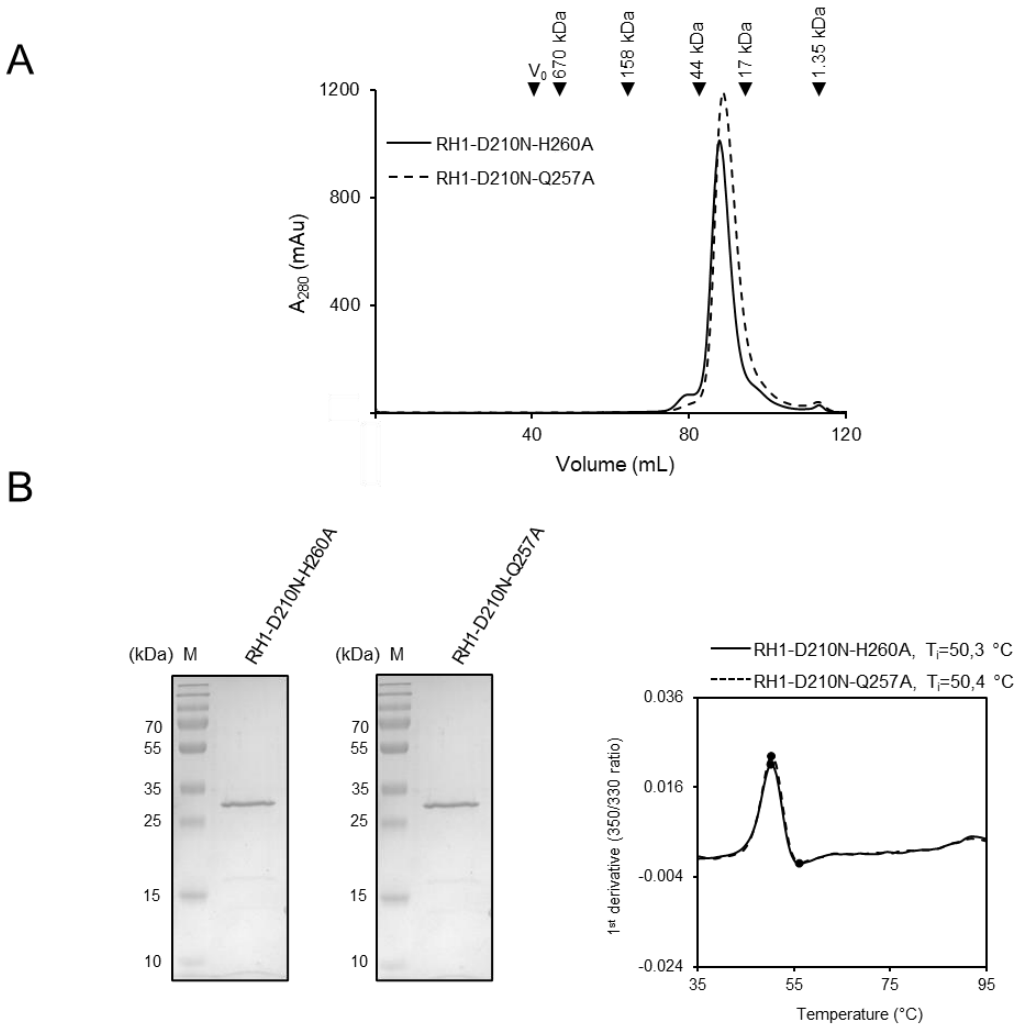


Figure 22. Prioritisation of EXOG-RNase H1 interfacial residues based on contact recurrence and interaction type. (A) Contact frequency heatmap based on the two highest-ranked AlphaFold2-Multimer models per EXOG-RNase H1 complex (one from multi-seed sampling and one from extended-recycling). Residue-level contacts were curated from AlphaFold2 predictions and PDBePISA analyses, following manual filtering of plausible inter-chain interactions. (B) Heatmap showing only inter-chain contacts identified exclusively by PDBePISA as hydrogen bonds or salt bridges. These contacts were extracted from the same two top-ranked AlphaFold2-Multimer models and represent chemically plausible and geometrically supported interactions, helping to define a confident subset of the interface.

7.1.7. Experimental validation of interface residues predicted by structure modelling

To assess the functional contribution of residues suspected to mediate the interaction between EXOG and RNase H1, five single-point alanine substitutions were introduced: three in EXOG (D169A, N331A, N334A) and two in RNase H1 (Q257A, H260A). All constructs were recombinantly expressed in *E. coli* BL21 (DE3) and purified under native conditions using affinity chromatography followed by size exclusion chromatography (SEC), according to procedures previously established for the wild-type proteins (Figure 23).



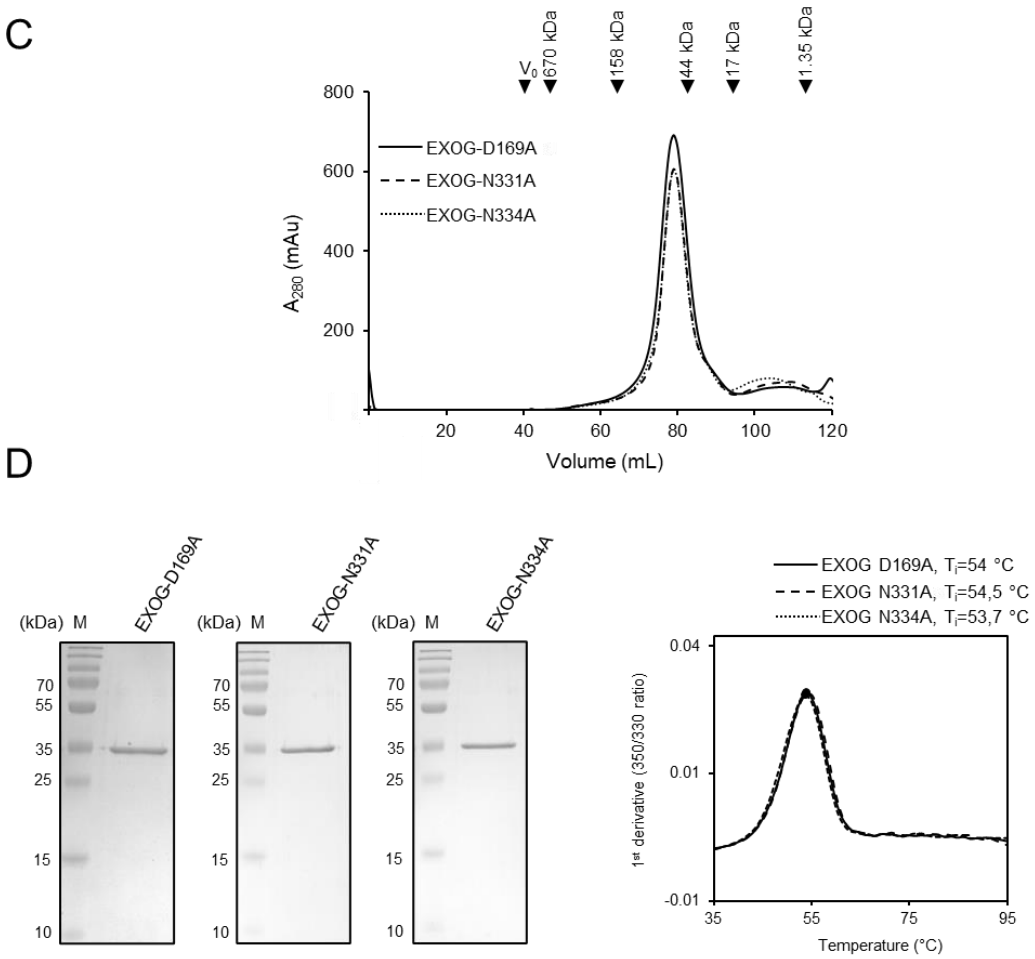


Figure 23. **Purification and characterisation of RNase H1 and EXOG variants bearing interface substitutions.** (A) Size exclusion chromatogram of RNase H1 variants separated on a HiLoad® 26/60 Superdex® 200 PG SEC column: RNase H1-D210N-H260A (black solid), RNase H1-D210N-Q257A (black dashed), (B) SDS0PAGE (15%) analysis of purified RNase H1 variants (1 μg per lane) stained with Coomassie Brilliant Blue. Molecular weight marker (M) is indicated. Thermal unfolding profiles of RNase H1 variants measured using Tycho NT.6. Both variants display similar inflection temperatures (T_i) around 50 °C, indicating preserved global fold. (C) Size exclusion chromatogram of EXOG variants separated on a HiLoad® 26/60 Superdex® 200 PG SEC column: EXOG D169A (black solid), EXOG N331A (black dashed), EXOG N334A (black dotted). (D) SDS-PAGE (15%) analysis of purified EXOG variants (1 μg per lane) stained with Coomassie Brilliant Blue. Molecular weight marker (M) is indicated. Thermal unfolding profiles of EXOG variants measured using Tycho NT.6. Tycho NT.6 analysis reveals uniform thermal unfolding transitions with Ti near 54 °C, indicative of proper folding and stability.

Size exclusion chromatography of all variants produced single, symmetric peaks at elution volumes expected for monomeric RNase H1 and for dimeric EXOG, consistent with preserved oligomeric states and absence of aggregation (Figure 23A and C). Coomassie stained SDS-PAGE of the purified samples showed single dominant bands at the expected apparent molecular masses with no visible degradation or major contaminants (Figure 23B and D, left panel). Thermal

stability measured with Tycho NT.6, which monitors changes in intrinsic tryptophan and tyrosine fluorescence during temperature ramps, yielded clear unfolding transitions with inflection temperatures characteristic for each protein in the tested buffer, indicating that alanine substitutions did not compromise global folding (Figure 23B and D, right panel). Together the chromatographic, electrophoretic and thermal signatures confirm successful overexpression and purification of all five proteins, retention of native assembly and solubility, and suitability of these interface mutants for the downstream binding and functional assays that follow.

To determine whether the selected single-residue substitutions affect complex formation, binding experiments were performed using two complementary biophysical techniques. Due to equipment availability, the RNase H1 mutants were assessed using SPR, while EXOG variants were analysed via BLI. All measurements were performed in HBS-EP buffer. For the SPR assay, wild-type EXOG was immobilised on a CM5 sensor chip, and serial dilutions of each RNase H1 mutant (D210N-Q257A and D210N-H260A) were injected as analytes. Sensorgrams for both variants displayed concentration-dependent association and dissociation profiles (Figure 24). Unexpectedly, both mutants retained robust binding activity and exhibited response levels comparable to or slightly exceeding those previously observed for the wild-type RNase H1. Although no kinetic fitting was performed due to the single replicate design, the binding curves suggested that the substitutions did not abolish interaction with EXOG under the tested conditions. In parallel, the binding of EXOG mutants (D169A, N331A, N334A) was evaluated using biotinylated RNase H1 immobilised on a streptavidin sensor in a BLI setup. Wild-type and mutant proteins were tested side by side at identical concentrations. All variants produced binding curves with similar association levels and signal intensities (Figure 24C), indicating that none of the substitutions significantly impaired binding under the experimental conditions applied.

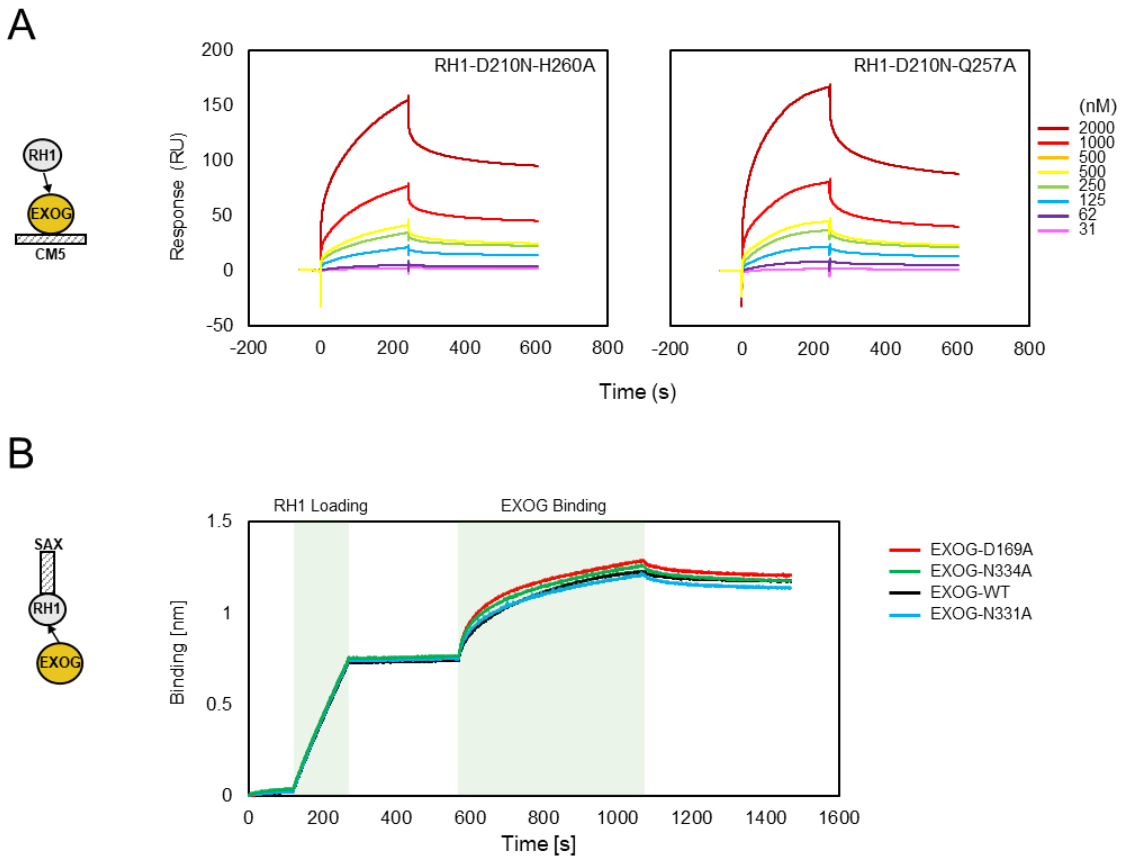


Figure 24. Binding analysis of RNase H1 and EXOG interface variants. (A) Surface plasmon resonance (SPR) sensorgrams showing the binding of RNase H1 variants D210N-H260A (A) and D210N-Q257A (B) to immobilised wild-type EXOG. Serial dilutions of each RNase H1 mutant (31.25 to 2000 nM) were injected over the sensor surface. Both variants exhibited concentration-dependent responses, consistent with retained binding activity. (B) Biolayer interferometry (BLI) binding curves comparing wild-type EXOG and interface variants D169A, N331A, and N334A, all tested at 1 µM concentration. Biotinylated wild-type RNase H1 (250 ng/mL) was immobilised on streptavidin biosensors. Following RNase H1 loading, sensors were transferred to wells containing EXOG for 150 seconds of association, followed by a wash. All steps were performed in 200 µL volumes using 96-well plates.

These findings suggest that the tested single-point mutations in EXOG and RNase H1 do not weaken the interaction between the two proteins. However, given the surprising preservation of binding despite substitutions at residues identified in structural analysis, it is possible that the interaction is either more redundant than initially hypothesised or that the current assay setup lacks the sensitivity required to detect subtle affinity differences. Further studies, potentially involving double or multiple substitutions, may be necessary to more definitively probe the functional relevance of individual interface residues.

7.1.8. Summary

All proteins required for this study were overexpressed, purified to homogeneity, and confirmed to be properly folded by SEC, CD, and melting analyses, which are documented in Figures 1 to 3. Surface plasmon resonance established domain dependent engagement of RNase H1 with EXOG, with full-length RNase H1 showing the tightest binding, Kd ≈ 0.38 μM; the catalytic domain alone retained substantial affinity, Kd ≈ 0.77 μM; inclusion of the native linker modestly improved the binding to the catalytic domain, Kd ≈ 0.50 μM; the hybridbinding domain alone bound weakly, Kd ≈ 12.6 μM, whereas adding the linker improved it to ≈ 1.43 μM (Figure 14). These affinity and rate trends indicate an interface centred on the catalytic domain, with the linker facilitating association primarily via increased on rates, and the hybrid-binding domain contributing only weakly on its own. Glycerol gradient co-sedimentation supported complex formation in solution, since co incubated proteins shifted to overlapping, heavier fractions relative to single protein controls, albeit by a modest amount consistent with the small mass contrast and the limited resolution of the gradient (Figure 15). A two-nuclease assay on a nicked 10RNA/10DNA hybrid showed that RNase H1 generates short RNA fragments at the junction that are then efficiently removed by EXOG, completing primer clearance; this cooperation was evident across reciprocal titrations and was reproduced with the disease variant V142I, which behaved comparably to wild-type, whereas A185V was impaired but could be partially complemented by EXOG (Figure 16 and 17). Crystallisation screens yielded one dataset at 2.3 Å in space group P2(1), where molecular replacement located only the EXOG dimer and no interpretable density for RNase H1, most likely due to flexibility of the full-length enzyme and crystal anisotropy (Figure 19 and Table 2). AlphaFold2 Multimer models consistently placed the catalytic domain at the core of the interface and achieved the highest interface confidence for the catalytic domain only complex, ipTM ≈ 0.650, with lower confidence for the full-length enzyme, ipTM ≈ 0.553, which is compatible with flexible linker and hybrid-binding domain regions that reduce model certainty (Figure 20). Contact mapping highlighted recurrent interfacial pairs that prioritised EXOG residues D169, N331, and N334 together with RNase H1 residues Q257, H260, and S265 for mutational testing, and the corresponding single alanine variants were well folded and assembled as expected; binding assays under the present formats showed preserved interaction, suggesting interface redundancy or limited sensitivity to small affinity changes, and motivating multiple substitutions and refined kinetics in future work (Figures 21 to 24).

7.2. Part II: Pol γ-mediated regulation of RNase H1 in RNA primer processing

In our previous study, we reported a direct interaction between RNase H1 and both the catalytic and accessory subunits of Pol γ, as demonstrated by microscale thermophoresis (MST) using purified proteins, and further validated by proximity ligation assay (PLA) in human cells (Karłowicz, 2022). Building on these observations, the cooperation between RNase H1 and Pol γ became one of the dedicated projects of my doctoral research, with the aim of exploring this interplay in greater depth and understanding its mechanistic basis.

7.2.1. Pol γ purification and structural integrity assessment

To enable structural and functional analyses of DNA polymerase γ (Pol γ) in the context of its newly identified interaction with RNase H1, the catalytic subunit Pol γA, its exonuclease-deficient variant Pol γA (exo-), and the accessory subunit Pol γB were overexpressed and purified to apparent homogeneity. Analytical size-exclusion chromatography revealed that Pol γA was monomeric under the tested conditions, eluting at a volume consistent with its predicted molecular weight (Figure 25A, dashed line). In contrast, Pol γB displayed an elution profile corresponding to a dimeric species (Figure 25A, dotted line). When the individually purified subunits (Pol γA (exo-) and Pol γB) were subsequently combined in a 1:1.1 molar ratio, the assembled Pol γ holoenzyme eluted as a single peak at a lower retention volume than either subunit alone (Figure 25A, solid line), consistent with complex formation. SDS-PAGE analysis confirmed the expected molecular weights of the individual subunits and the holoenzyme, and verified the absence of major contaminants or degradation products (Figure 25A, right panel).

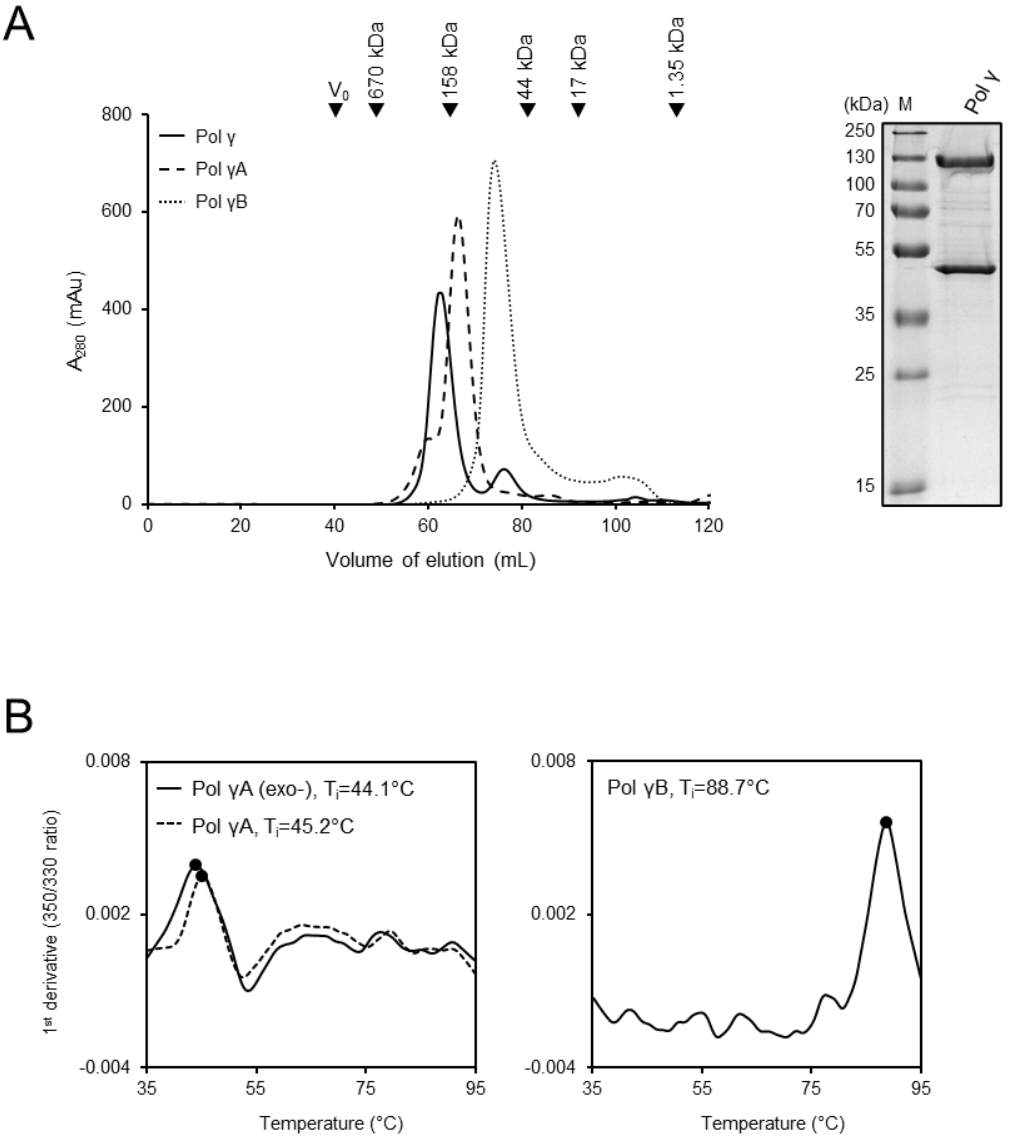


Figure 25. Purification and characterisation of Pol γA, Pol γB. (A) Size-exclusion chromatogram of Pol γA (dashed line), Pol γB (dotted line), and the Pol γ holoenzyme (solid line), separated on a HiLoad Superdex 200 pg preparative column. Right: SDS-PAGE analysis of the same protein samples (1 µg each) stained with Coomassie Brilliant Blue and resolved on a 12.5% gel. Molecular weight marker (M) is indicated. (B) Thermal stability of the proteins was assessed using intrinsic fluorescence measurements on a Tycho NT instrument. The first derivative of the fluorescence ratio at 350 nm/330 nm is shown. Pol γA (exo-) is represented by a solid line, and wild-type Pol γA by a dashed line. Dots indicate inflection points (T_m), with corresponding values shown on the graph.

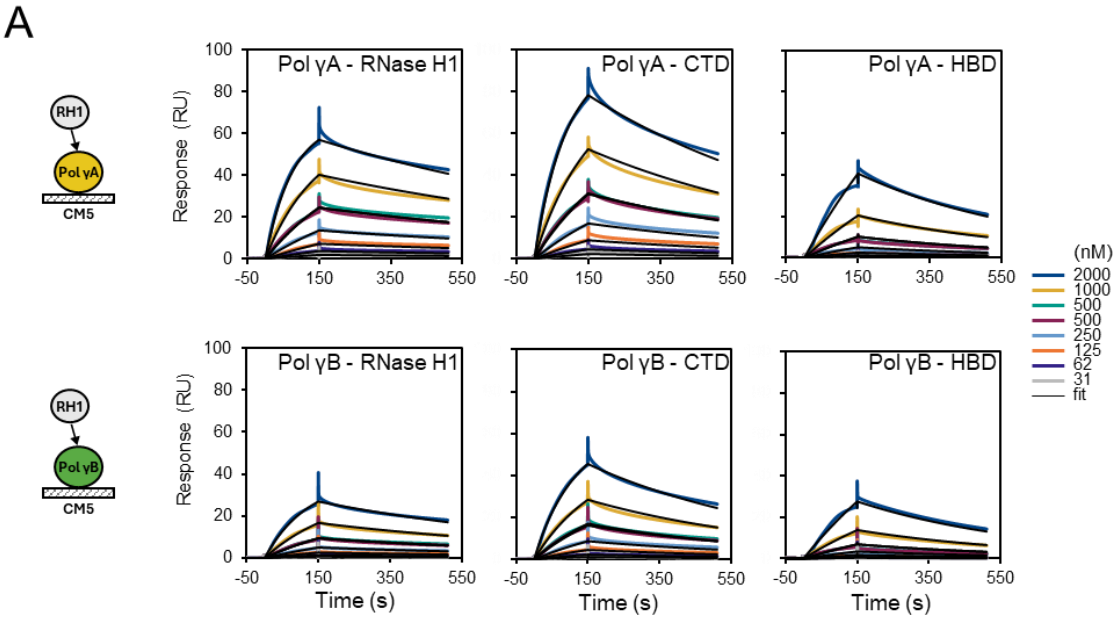
The structural integrity and folding properties of the Pol γ subunits were assessed by thermal unfolding experiments using intrinsic tryptophan/tyrosine fluorescence on a Tycho NT instrument. The unfolding transitions were well-defined for all proteins, allowing determination of the inflection temperature (Ti) from the maximum of the first derivative of the fluorescence ratio at 350/330 nm (Figure 25B). Wild-type Pol γA exhibited a Ti of 45.2 °C, while the

exonuclease-deficient variant showed a slightly reduced stability with a Ti of 44.1 °C. In contrast, Pol γB was markedly more thermostable, unfolding at a Ti of 88.7 °C. These Ti values are consistent with previously reported data for human Pol γ subunits (Dosekova et al., 2020; Silva-Pinheiro et al., 2021), confirming the high stability of the accessory subunit relative to the catalytic subunit. Together, these analyses verified that all Pol γ subunits and the holoenzyme are properly folded, homogeneous, and biochemically stable under the tested conditions, thereby establishing their suitability for subsequent interaction studies with RNase H1 and functional assays (Figure 25).

7.2.2. Catalytic domain of RNase H1 mediates direct interaction with Pol γ subunits

To determine which regions of RNase H1 contribute to binding with the human mitochondrial DNA polymerase γ subunits, I conducted surface plasmon resonance (SPR) experiments using a CM5 sensor chip. Both Pol γA (the catalytic subunit) and Pol γB (the accessory subunit) were immobilised in separate flow channels via standard amine coupling. Increasing concentrations of three RNase H1 constructs were injected as analytes: full-length RNase H1, CTD, and HBD. Each analyte was titrated in a series of concentrations ranging from 31 nM to 2 μM, and binding curves were obtained in at least two independent experiments per construct. Reference-subtracted data were fitted globally to a 1:1 Langmuir binding model to extract kinetic parameters (Figure 26).

A



B

Ligand	Analyte	k_{on} (10^3 $M^{-1}s^{-1}$)	k_{off} (10^{-3} s^{-1})	K_d (10^{-6} M)
Pol γA	RNase H1	6.20 ± 0.77	0.93 ± 0.12	0.15 ± 0.03
	CTD	5.16 ± 1.41	1.41 ± 0.12	0.29 ± 0.07
	HBD	0.07 ± 0.01	1.96 ± 0.38	29 ± 5
Pol γB	RNase H1	3.64 ± 0.74	1.28 ± 0.13	0.36 ± 0.08
	CTD	3.85 ± 1.41	1.73 ± 0.22	0.63 ± 0.04
	HBD	0.03 ± 0.01	2.06 ± 0.03	139 ± 82

Figure 26. Surface plasmon resonance reveals domain-dependent interactions between RNase H1 and EXOG. (A) SPR sensorgrams showing the binding of RNase H1 variants to immobilised Pol γA (top) and Pol γB (bottom). Increasing concentrations of full-length RNase H1, CTD, and HBD (31.25, 62.5, 125, 250, 500, 1000, and 2000 nM) were injected over the sensor chip surface, and binding responses were recorded. (B) Kinetic parameters derived from the interactions shown in (A). Association rate constants (k_{on}), dissociation rate constants (k_{off}), and equilibrium dissociation constants (K_d) were determined using a 1:1 binding model. Values represent the mean ± standard deviation from at least two independent experiments with global fitting.

Full-length RNase H1 exhibited robust binding to both Pol γ subunits, with dissociation constants (K_d) of 0.15 ± 0.03 µM for Pol γA and 0.36 ± 0.08 µM for Pol γB. The association rate constants (k_{on}) were 2.66 ± 0.13 × 10^3 M^{-1} s^{-1} for Pol γA and 1.42 ± 0.12 × 10^3 M^{-1} s^{-1} for Pol γB, while the corresponding dissociation rate constants (k_{off}) were 0.40 ± 0.02 × 10^{-3} s-1 and 0.51 ± 0.03 × 10^{-3} s^{-1}, respectively. When tested in isolation, the CTD retained measurable affinity for both subunits, with K_d values of 0.29 ± 0.07 µM for Pol γA and 0.63 ± 0.04 µM for Pol γB. The k_{on} values (1.85 ± 0.09 × 10^3 M^{-1} s^{-1} for Pol γA; 1.37 ± 0.08 × 10^3 M^{-1} s^{-1} for Pol γB) and k_{off} values (0.54 ± 0.03 × 10^{-3} s^{-1} and 0.86 ± 0.06 × 10^{-3} s^{-1}, respectively) were comparable to those

obtained for the full-length protein, indicating that this domain alone can establish a stable complex with both Pol γA and Pol γB. In contrast, the HBD displayed markedly weaker binding to both subunits. The measured affinities were 29 ± 5 μM for Pol γA and 139 ± 82 μM for Pol γB, with very low k_{on} values (0.06 ± 0.02 × 10^3 M^{-1} s^{-1} for Pol γA; 0.03 ± 0.01 × 10^3 M^{-1} s^{-1} for Pol γB) and correspondingly higher k_{off} values (1.76 ± 0.07 × 10^{-3} s^{-1} and 4.35 ± 0.31 × 10^{-3} $s^{-1,}$ respectively). These kinetic parameters are consistent with transient or weak interactions for this domain in isolation. Together, our binding results indicate that the CTD domain of RNase H1 is responsible for the direct interaction with both Pol γA and Pol γB subunits, while the HBD domain seems to have a negligible impact on the protein-protein complex formation.

7.2.3. Catalytically inactive RNase H1 hinders gap-filling by Pol γ

To investigate the functional relationship between RNase H1 and Pol γ, I first examined the activity of Pol γ alone on a gapped RNA/DNA substrate (Figure 27A). Within the first minute of the reaction, nucleotide incorporation by Pol γ was limited, as indicated by a ladder of intermediate-length products on the gel corresponding to partially gap-filling species (Figure 27B, lanes 2-5). After 10 minutes, a prominent band corresponding to the complete 10-nt gap-filling product was observed (Figure 27B, lanes 6-8). The reaction was performed with an excess of substrate (50 nM) relative to Pol γ (12.5 nM) to minimise excessive downstream RNA displacement, and to ensure that observed products reflect gap-filling rather than strand displacement synthesis.

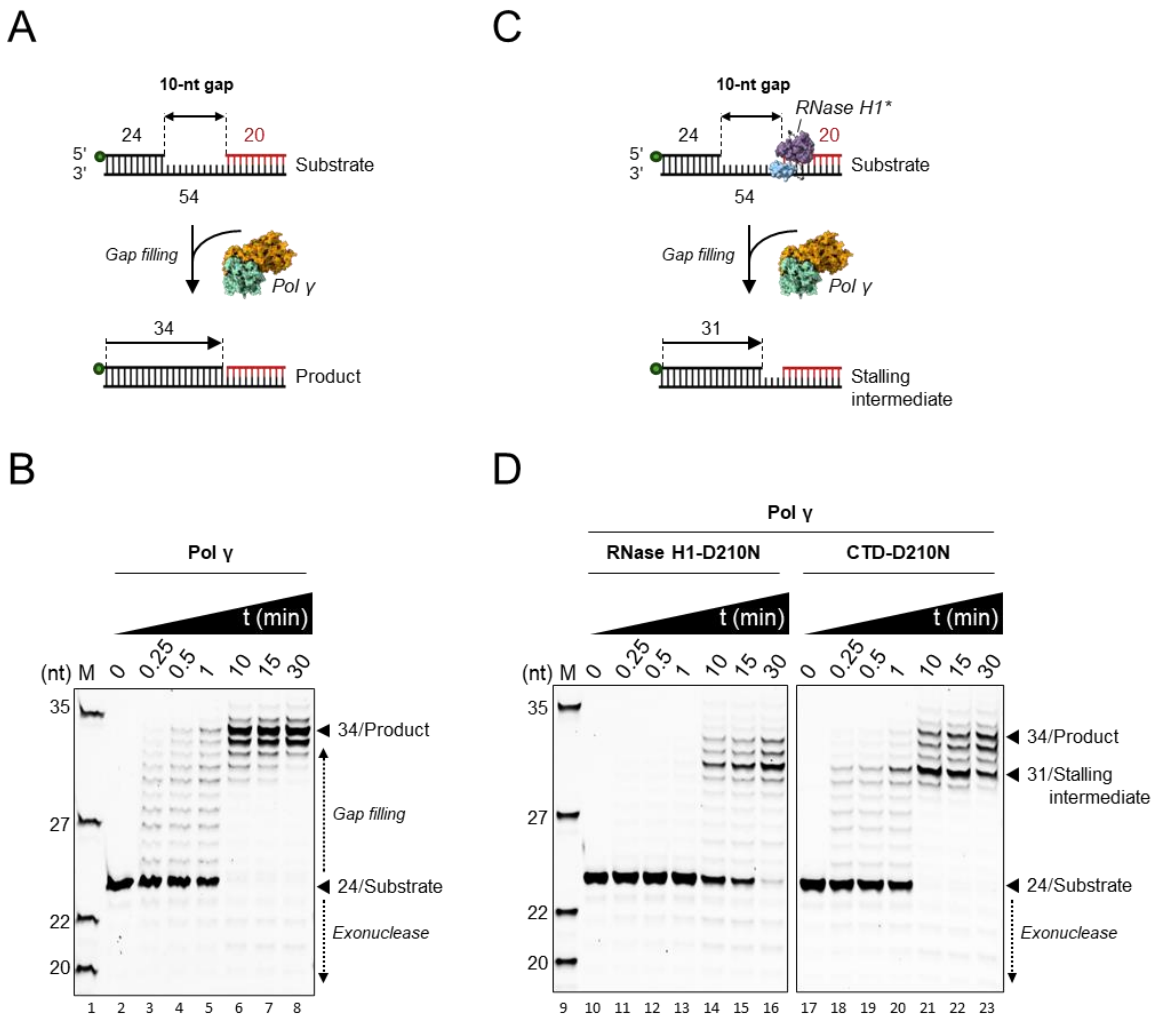


Figure 27. RNase H1 creates a barrier to Pol γ-mediated DNA synthesis. (A) Schematic representation of the gap-filling reaction shown in **(B)**. 50 nM of gapped RNA/DNA substrate, labelled at the 5′ end with fluorescein, was incubated with 12.5 nM Pol γ for 0, 0.25, 0.5, 1, 10, 15, and 30 min at 25 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 1 μM dNTPs. At the indicated time points, reactions were stopped and analysed by 20% denaturing Urea-PAGE. (B) Denaturing gel electrophoresis analysis of Pol γ gap-filling activity from reactions described in (A). M - molecular weight standard. (C) Schematic representation of the gap-filling reaction in the presence of RH1 shown in (D). The 5′ end of the substrate strand is labelled with fluorescein. 50 nM of gapped RNA/DNA substrate was incubated with 12.5 nM Pol γ in the presence of 1.5 μM RNase H1-D210N or CTD-D210N for 0, 0.25, 0.5, 1, 10, 15, and 30 min at 25 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 1 μM dNTPs. At the indicated time points, reactions were stopped and analysed by 20% denaturing Urea-PAGE. (D) Denaturing gel electrophoresis analysis of Pol γ activity in the presence of RNase H1, from reactions described in (C). The left gel corresponds to RNase H1-D210N, while the right gel corresponds to CTD-D210N. M - molecular weight standard.

When the catalytically inactive RNase H1-D210N was included (Figure 16C), a distinct accumulation of a 31-nt intermediate was detected, consistent with stalling of Pol γ after the

incorporation of 7 nucleotides (Figure 27D, lanes 14-16). No detectable DNA synthesis occurred within the first minute (Figure 27D, lanes 10-13), suggesting that RNase H1-D210N may interfere with initial polymerase binding or elongation on the substrate. The exonuclease activity of Pol γ, although limited under these reaction conditions, was unaffected by RNase H1-D210N (compare Figure 27B, lanes 3-8, with Figure 27D, lanes 11-16). These findings indicate that the polymerase activity of Pol γ, but not its exonuclease activity, is impeded in the presence of catalytically deficient RNase H1. Additionally, I performed the experiment using the isolated CTD carrying the D210N mutation to assess whether the stalling effect was specific to full-length RNase H1. Previous studies have shown that the hybrid-binding domain (HBD) of eukaryotic RNase H1 greatly enhances enzyme processivity, and that its removal results in distributive cleavage similar to bacterial RNase HI (Gaidamakov et al., 2005; Nowotny et al., 2008) . I hypothesised that if the stalling of Pol γ by RNase H1-D210N results from stable retention of the nuclease on the RNA/DNA hybrid, then the CTD-D210N variant, having reduced substrate affinity, should exhibit diminished stalling. Indeed, in the presence of CTD-D210N, Pol γ displayed polymerase activity within the first minute of the reaction, initially pausing after incorporation of 7 nucleotides (Figure 27D, lanes 18-20). Over the course of the reaction, Pol γ was able to fully fill the gap, indicating that the initial pause was transient (Figure 27D, lanes 21-23). These results support the conclusion that the HBD domain of RNase H1 facilitates prolonged retention of the catalytically inactive mutant on the RNA/DNA substrate, thereby impeding Pol γ progression and blocking complete DNA gap-filling.

7.2.4. RNase H1 retention on RNA/DNA hybrid does not induce Pol γ dissociation from the gap

To determine whether the observed incomplete gap-filling by Pol γ in the presence of catalytically inactive RNase H1 could result from premature dissociation of Pol γ from the substrate, I employed biolayer interferometry (Sultana & Lee, 2015) to simultaneously monitor the binding dynamics of Pol γ and RNase H1 to the same substrate. Using streptavidin-coated biosensors, I immobilised biotinylated RNA/DNA hybrids (step 1) and then sequentially introduced Pol γ (step 2) and RNase H1-D210N (step 3), monitoring all binding events in real time (Figure 28, top scheme).

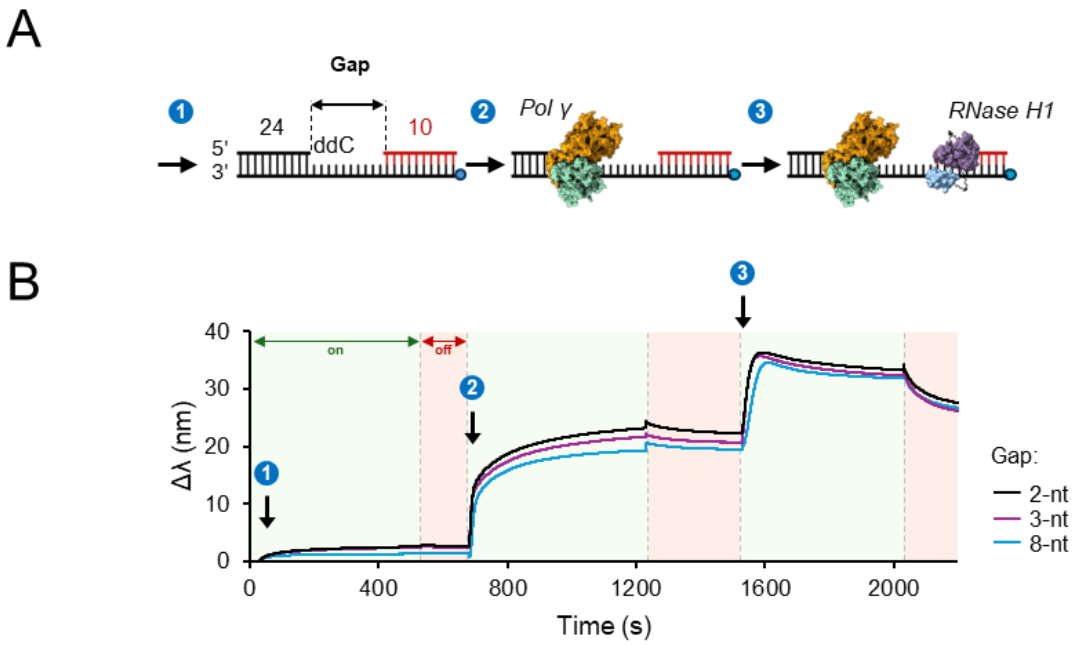


Figure 28. Sequential binding of Pol γ and catalytically inactive RNase H1 to a gapped RNA/DNA hybrid monitored by BLI. (A) Schematic representation of Pol γ and RNase H1 complex formation. (1) biotinylated RNA/DNA substrate. (2) Pol γ binding to the substrate. (3) RNase H1 accommodating on the same substrate with Pol γ. (B) Representative BLI sensograms showing association ("on") and dissociation ("off") phases for each binding step. In the experiment, 1 μM of gapped RNA/DNA substrate was incubated with the SAX biosensor, washed, incubated with 1 μM Pol γ, washed again, and finally incubated with 1 μM RNase H1-D210N. At each stage, a distinct increase in signal was observed, consistent with sequential protein loading and stable complex formation. The accompanying legend indicates the tested substrates containing gaps of 2, 3, or 8 nucleotides.

To assess whether gap length influences the binding dynamics of these enzymes, I tested substrates containing gaps of 2 and 3 nt (corresponding to the stalling intermediates observed in my gap-filling experiments; see Figure 27D) as well as an 8 nt gap. The primer strand contained a 3′ dideoxycytidine (ddC) to prevent nucleotide incorporation. Upon binding of Pol γ, an increase in signal was observed during the association phase ("on"), followed by a slow dissociation ("off") for all substrates. Subsequent introduction of RNase H1-D210N produced a further increase in signal, consistent with its binding to the substrate. A moderate reduction in signal during the association phase was observed for RNase H1-D210N across all substrates tested. This decrease could arise from slight destabilisation of the interaction between Pol γ and/or RNase H1-D210N with the substrate, as well as the inherent instability of the RNase H1 variant under reaction conditions.

To examine whether the order of protein addition affected binding, I repeated the experiment with the 2 nt gap substrate but introduced RNase H1-D210N first (Figure 29A). In this reversed

order, initial binding of RNase H1-D210N resulted in a single-phase increase in signal, followed by a plateau during the association phase. Subsequent addition of Pol γ caused a further increase in signal, indicating its binding to the substrate, followed by slow dissociation. The association and dissociation profiles for each enzyme were comparable to those observed in the main experiment, indicating that binding behaviour was independent of the order of protein addition.

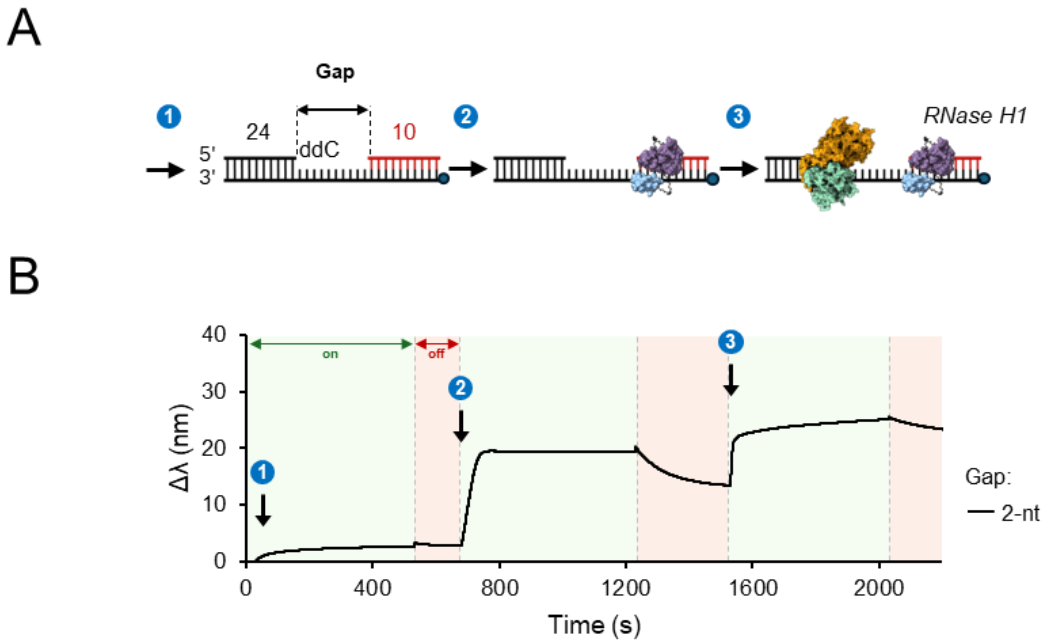


Figure 29. Reversed-order experiment of RNase H1-D210N and Pol γ binding to a gapped RNA/DNA hybrid. (A) Schematic representation of Pol γ and RNase H1 complex formation. (1) biotinylated RNA/DNA substrate. (2) Binding of 1 μM RNase H1-D210N to the substrate.(3) Binding of 1 μM Pol γ holoenzyme to the same substrate already occupied by RNase H1-D210N. (B) Representative BLI sensograms showing association ("on") and dissociation ("off") phases for each binding step. In the experiment, 1 μM of gapped RNA/DNA substrate was incubated with the SAX biosensor, washed, incubated with 1 μM RNase H1-D210N, washed again, and finally incubated with 1 μM Pol γ. Each binding step produced a clear increase in signal, indicating sequential loading and stable complex formation. The accompanying legend indicates the tested substrate containing gap of 2 nucleotides.

Additionally, a control experiment without immobilised substrate confirmed the absence of nonspecific adsorption of either protein to the sensor surface (Figure 30).

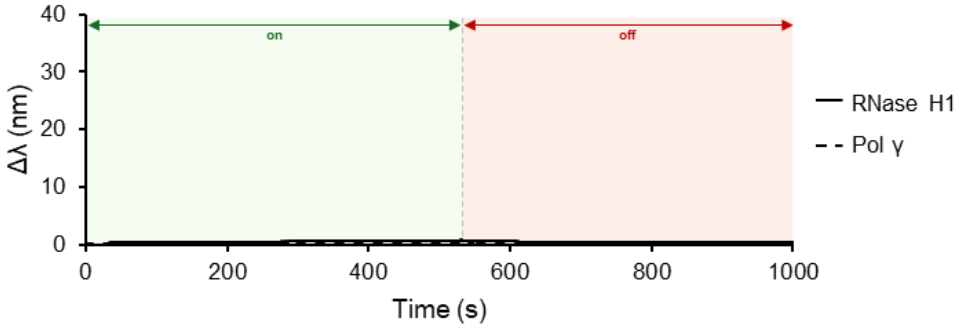


Figure 30. **Control experiments for BLI-based ternary complex formation.** Control experiments demonstrating the absence of nonspecific binding of Pol γ or RNase H1-D210N to the biosensor surface. SAX biosensors without immobilised substrate were incubated with either RNase H1-D210N alone (solid line) or Pol γ alone (dashed line), under conditions identical to those used in Figure 2E. No detectable increase in BLI signal was observed.

Together, these BLI results demonstrate that Pol γ and RNase H1-D210N can stably bind a common RNA/DNA hybrid substrate, and that their binding affinity is unaffected by gap length or by the order in which the proteins are added. We therefore conclude that the stalling of Pol γ observed in the presence of catalytically inactive RNase H1 is not caused by Pol γ dissociation, but rather by retention of RNase H1 on the RNA/DNA hybrid, which creates a physical barrier that Pol γ cannot bypass.

7.2.5. Pol γ stimulates RNase H1-mediated RNA degradation, facilitating complete RNA removal

Building on the observation that catalytically deficient RNase H1 impairs the polymerase activity of Pol γ without affecting its DNA binding, I examined whether wild-type RNase H1 activity is reciprocally influenced by the presence and catalytic function of Pol γ. To this end, I used a substrate containing a 10-nt gap to enable Pol γ-driven DNA synthesis in the presence of dNTPs, followed by a downstream RNA/DNA hybrid for RNase H1 binding and cleavage (Figure 31A).

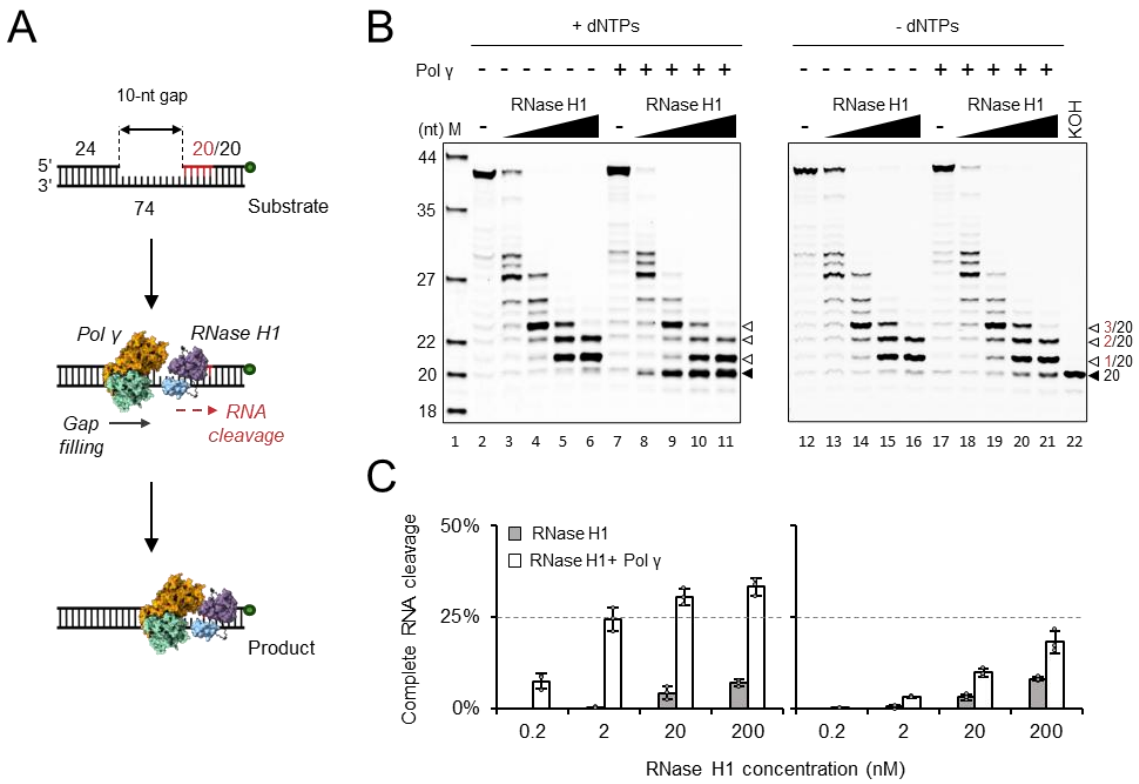


Figure 31. Pol γA enhances RNase H1 activity, enabling complete RNA primer digestion. (A) Schematic representation of the RNase H1 nuclease assay. 200 nM gapped RNA/DNA substrate, labelled at the 3′ end with fluorescein, was incubated with increasing concentrations of RNase H1 (0.2, 2, 20, 200 nM) and with or without 100 nM Pol γ for 1 h at 37 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 100 μM dNTPs (where indicated), then stopped and analysed by 20% denaturing Urea-PAGE. (B) Denaturing gel electrophoresis analysis of RNase H1 activity with and without Pol γ. The left gel corresponds to reactions conducted in the presence of dNTPs, while the right gel corresponds to reactions conducted in the absence of dNTPs. M - molecular weight standard. (C) Densitometric analysis of cleavage products, with the graph representing mean values ± standard deviation (error bars) from three independent experiments.

Adjacent to the hybrid region, a DNA duplex with a fluorescein-labelled 3′ end, which is not cleaved by RNase H1, enabled tracking of RNA processing. In control reactions with RNase H1 alone, the expected pattern of partial RNA degradation was observed, leaving 1, 2 or 3 ribonucleotides at the 5′ end of the DNA strand (Figure 31B, lanes 2-6, open arrowheads). Strikingly, addition of Pol γ led to an accumulation of fully cleaved, RNA-free product (Figure 31B, lanes 7-11, black arrowhead). The intensity of this band increased with higher RNase H1 concentration, indicating that complete RNA removal in the presence of Pol γ scales with the amount of RNase H1 (Figure 31C, left graph). To test whether the accessory subunit of Pol γ is sufficient to stimulate RNA removal by RNase H1, I repeated the experiment in the presence of RNase H1 and Pol γB only (Figure 32A). No stimulation of RNase H1-mediated RNA cleavage

was detected under these conditions (Figure 32B), showing that Pol γB alone does not promote complete RNA removal.

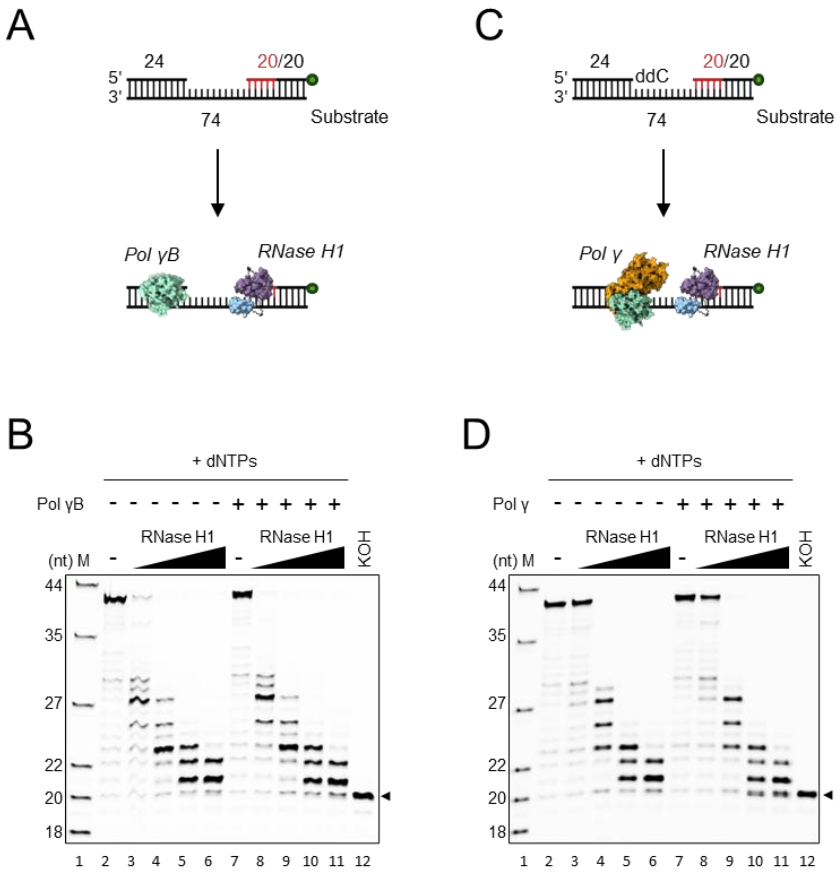


Figure 32. Pol γ catalytic activity may be required for RNase H1-mediated RNA primer cleavage. (A) Schematic representation of the RNase H1 nuclease assay performed in the presence of 100 nM Pol γB. A 200 nM gapped RNA/DNA substrate labelled at the 3′ end with fluorescein was incubated with increasing concentrations of RNase H1 (0.2, 2, 20, 200 nM) for 1 h at 37 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 100 μM dNTPs. (B) Denaturing gel electrophoresis analysis of RNase H1 activity with and without Pol γB, from reactions described in (A). M - molecular weight marker. (C) Schematic representation of the RNase H1 nuclease assay using a ddC-blocked RNA/DNA substrate. 200 nM gapped RNA/DNA substrate, labelled at the 3′ end with fluorescein, contained a ddC-terminated primer to prevent Pol γ-mediated polymerisation. Reactions were performed with increasing concentrations of RNase H1 (0.2, 2, 20, 200 nM), in the presence or absence of 100 nM Pol γ and incubated for 1 h at 37 °C with 10 mM $MgCl_2$ and 100 μM dNTPs. (D) Denaturing gel electrophoresis analysis of RNase H1 activity with and without Pol γ under ddC-blocked conditions, from reactions described in (C). M - molecular weight marker.

To assess whether complete RNA removal depends on gap-filling by Pol γ-mediated nucleotide incorporation, I omitted dNTPs. Under these conditions, fully cleaved product was strongly reduced yet still detectable (Figure 31B, lanes 17-21; Figure 31C, right graph). dNTPs had no effect on RNase H1 activity in the absence of Pol γ (Figure 31B, compare lanes 2-6 with 12-16).

I next used a ddC-blocked primer to prevent strand extension while keeping dNTPs present (Figure 32C). As in the dNTP-omitted reactions, there was no detectable stimulation of complete RNA removal by RNase H1 (Figure 32D). Together, these data show that gap-filling by Pol γ markedly enhances RNase H1-dependent complete RNA cleavage. Whether the enhancement arises from DNA synthesis per se or from the physical proximity of both enzymes on the same substrate remains to be clarified.

Next, I asked whether Pol γ also stimulates the isolated catalytic domain of RNase H1. Under identical conditions, the CTD alone displayed lower cleavage efficiency on the RNA/DNA hybrid than the full-length enzyme (Figure 33A, lanes 2-6; compare with Figure 31B, lanes 2-6). Upon addition of Pol γ, a fully cleaved product was again observed (Figure 33A, lanes 7-11, black arrowhead), although its intensity was reduced relative to the full-length RNase H1 (Figure 33B, left graph; compare with Figure 31C, left graph), consistent with the overall reduced activity of the truncated variant. Omitting dNTPs reduced complete RNA primer removal (Figure 33A, lanes 12-21; Figure 33B, right panel), indicating that Pol γ-driven gap-filling promotes this step.

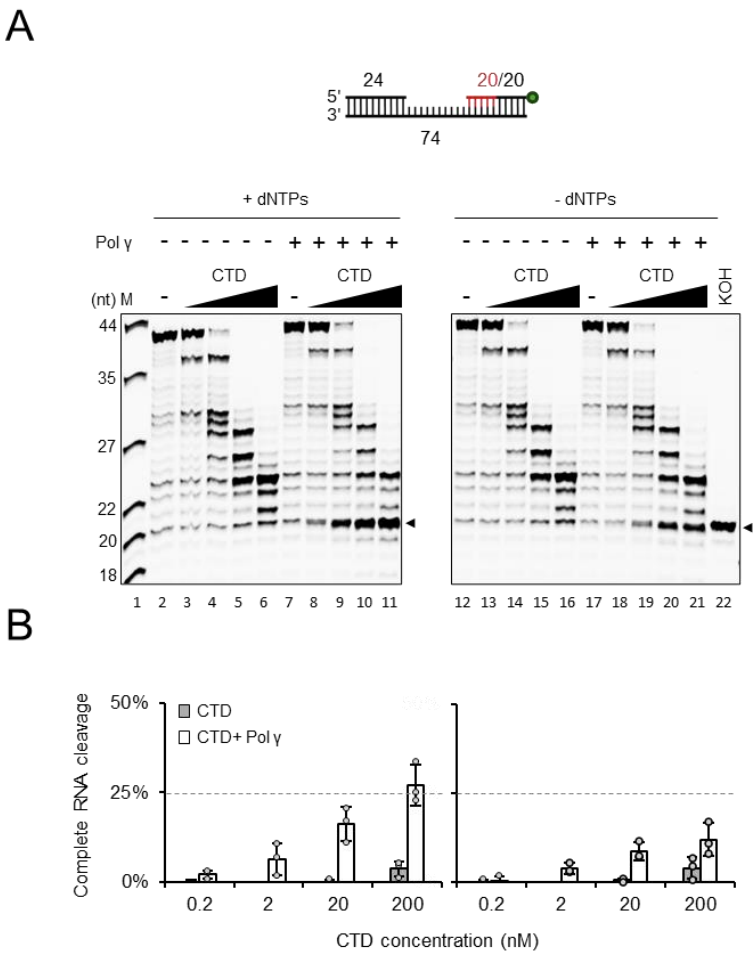


Figure 33. Pol γ stimulates the catalytic activity of the RNase H1 CTD variant. (A) Denaturing gel electrophoresis analysis of RNA cleavage by RNase H1 CTD. 200 nM gapped RNA/DNA substrate, labelled at the 3′ end with fluorescein, was incubated with increasing concentrations of RNase H1 CTD (0.2, 2, 20, 200 nM) in the presence or absence of 100 nM Pol γ for 1 h at 37 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 100 μM dNTPs (where indicated), then analysed by 20% denaturing Urea-PAGE. (B) Densitometric analysis of cleavage products. Graphs represent mean ± standard deviation from three independent experiments.

Finally, I examined whether the length of the RNA tract influences Pol γ-dependent stimulation of RNase H1. Based on previous studies indicating that human RNase H1 requires a minimal RNA/DNA duplex length of six base pairs (Wu et al., 1999), I tested substrates containing four or two consecutive RNA nucleotides (Figure 34A and 34C). RNase H1 alone processed the 4-RNA substrate, albeit less efficiently than the 20-RNA substrate (Figure 34A, lanes 2-6 and 12-16; compare with Figure 31B, lanes 2-6 and 12-16). In contrast, no activity was detected on the 2-RNA substrate (Figure 34C).

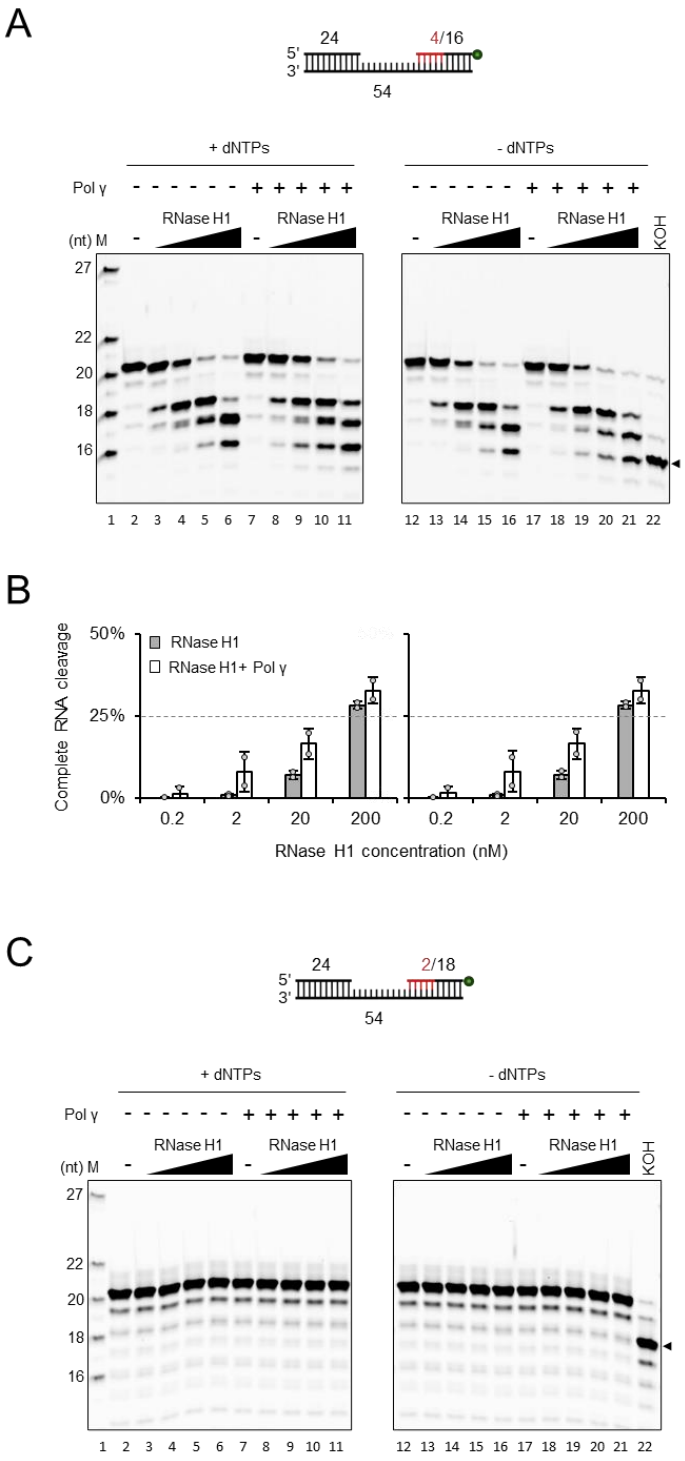


Figure 34. Pol γ stimulates RNase H1 activity on RNA/DNA hybrids depending on RNA length. (A, C) Denaturing gel electrophoresis analysis of RNase H1 activity on substrates containing either two (A) or four (B) consecutive RNA nucleotides. 200 nM RNA/DNA substrates, labelled at the 3′ end with fluorescein, were incubated with increasing concentrations of RNase H1 (0.2, 2, 20, 200 nM) with or without 100 nM Pol γ for 1 h at 37 °C. Reactions were supplemented with 10 mM $MgCl_2$ and 100 μM dNTPs (where indicated), then analysed by 20% denaturing Urea-PAGE. (B) Densitometric analysis of cleavage products. Graphs represent mean ± standard deviation from two independent experiments.

7.2.6. PolγAB and RNase H1 form a stable higher-order complex with the RNA/DNA hybrid

To assess the formation of a stable Pol γ-RNaseH1-RNA/DNA hybrid (S7, Table 1) complex, size exclusion chromatography (SEC) was performed, comparing the elution profiles of Pol γ alone, RNase H1 alone, and the Pol γ-RNase H1-substrate mixture. The control run of Pol γ alone (orange line) resulted in a single major peak at a defined elution volume, corresponding to the molecular weight of the Pol γ complex (Figure 35A). Similarly, the control run of RNase H1 alone (purple line) produced a distinct single peak at a later retention volume, consistent with its lower molecular weight.

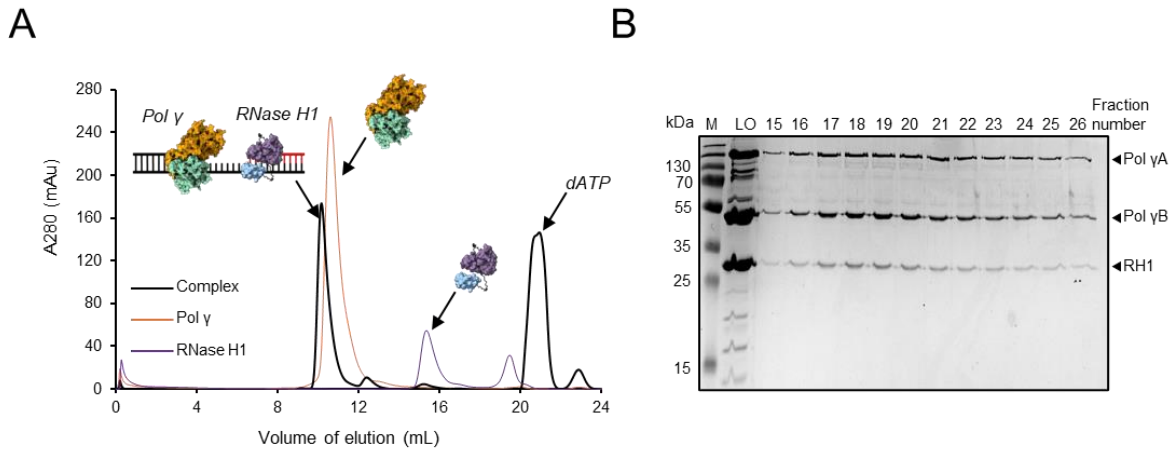


Figure 35. Pol γ and RNase H1 form a stable ternary complex with an RNA/DNA hybrid observed on SEC. (A) Size exclusion chromatography (SEC) profiles showing the elution of Pol γAB alone (orange), RNase H1 alone (purple), and a mixture of Pol γAB, RNase H1, and RNA/DNA hybrid substrate (black). The mixture contained 2.5 µM Pol γAB, 7.5 µM RNase H1, and 2.5 µM RNA/DNA substrate (molar ratio 1:3:1). (B) SDS-PAGE of fractions 16-25 collected from the main elution peak in panel A. Bands corresponding to Pol γA, Pol γB, and RNase H1 are indicated.

For the Pol γ-RNase H1-substrate mixture (black line), the components were combined at the following final concentrations: 2.5 µM Pol γ, 7.5 µM RNase H1, and 2.5 µM RNA/DNA hybrid substrate, resulting in a molar ratio of 1:3:1 (Pol γ:RNase H1:substrate). The SEC profile of this mixture showed a clear shift in the main elution peak to a lower retention volume compared to Pol γ alone, indicating an increase in molecular mass and suggesting the formation of a higher-order complex containing all components. Importantly, despite RNase H1 being in excess (3:1 relative to Pol γ and the substrate), its previously observed peak from the control run was no longer predominant, suggesting that the majority of RNase H1 was bound to the complex rather than remaining unbound in solution. To confirm that all three proteins co-eluted in the shifted

peak, fractions corresponding to the main elution peak (fractions 16-25) were analysed by SDS-PAGE (Figure 35B). The gel revealed co-elution of Pol γA, Pol γB, and RNase H1, confirming that all three proteins were present in the same fractions. The lack of RNaseH1 in later fractions further supports its binding to the complex rather than remaining free in solution. These results demonstrate that Pol γ and RNase H1 form a stable complex with the RNA/DNA hybrid substrate, as evidenced by the SEC elution shift and co-elution of all three proteins in SDS-PAGE.

7.2.7. Cryo-EM strategy for the Pol γ-RNase H1-RNA/DNA complex

The biochemical and biophysical analyses described in Part II demonstrated that Pol γ and RNase H1 form a stable ternary complex with an RNA/DNA hybrid, and that their functional interplay is relevant for complete RNA primer removal. These findings established a solid experimental foundation to investigate the architecture of the complex at high resolution. Single-particle cryo-electron microscopy (cryo-EM) was therefore chosen as the structural method, offering the potential to visualise the entire assembly in a near-native state without the constraints of crystallisation. In designing the cryo-EM study, particular emphasis was placed on sample composition, stoichiometry, and biochemical validation, to ensure that the specimens used for grid preparation reflected a functionally competent state of the complex. This strategy was intended to maximise the likelihood of capturing the relevant structural features and to provide a framework for interpreting potential conformational variability. Across this project I undertook a broad and systematic effort to obtain a cryo-EM competent Pol γ-RNase H1-RNA/DNA substrate assembly. Gap length was varied from two to ten nucleotides and multiple hybrid designs with different RNA and DNA tract lengths were tested. Complexes were assembled with full-length RNase H1 and with the isolated CTD, alternative orders of component addition were explored, and a wide range of molar ratios between Pol γ and RNase H1 was screened. Isolation of the complex was attempted by size-exclusion chromatography and by glycerol gradient centrifugation, with or without crosslinkers added during assembly or after purification. Multiple grid supports were screened, including gold and graphene, and vitrification parameters in the Vitrobot were optimised. Despite this breadth of strategies, outcomes on grids or after data processing consistently fell into four categories: aggregation, absence of particles, dissociation of the complex, or apparently promising screening that nonetheless never produced interpretable RNase H1 density. In view of this extensive yet inconclusive exploration, I present in detail only the two processing workflows that were most promising. They yielded high-quality reconstructions of Pol γ on DNA substrates but did not provide an averaged RNase H1 volume and therefore did not enable solution of the complete complex structure.

Complex 1

A ternary complex containing the Pol γ holoenzyme, catalytically inactive full-length RNase H1 D210N, and an RNA/DNA hybrid substrate was selected for single particle cryo-EM analysis. The substrate consisted of a 24 nt primer-template duplex terminated with ddC at the 3′ end, followed by a 5 nt single stranded DNA gap and a downstream duplex of 20 nt in which the first 10 nt were RNA and the remaining 10 nt DNA (S10, Table 1 and Figure 36). The complex was assembled at final concentrations of 2 µM Pol γ, 5 µM RNase H1, and 2 µM substrate, and purified by glycerol gradient ultracentrifugation. Fractions showing co-migration of Pol γ and RNase H1 were pooled and concentrated, yielding a homogeneous preparation suitable for cryo-EM (Figure 36). Samples were vitrified and data were collected as described in Materials and Methods.

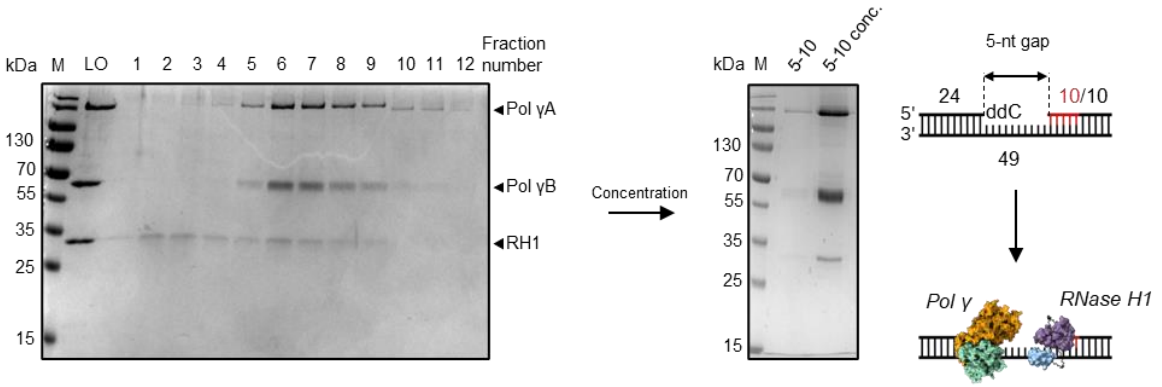


Figure 36. **Glycerol gradient ultracentrifugation reveals co migration of Pol γ and RNase H1 on RNA/DNA substrate**. From left to right, SDS-PAGE of twelve fractions collected from the top of a 2% to 15% glycerol gradient shows co migration of Pol γA and Pol γB with RNase H1. M, molecular mass marker; LO, load. Fractions 5 to 10 were pooled, concentrated, and used for cryo-EM grid preparation. Next, SDS-PAGE of the pooled concentrate is shown. At right, a schematic representation of complex formation within the gradient.

Image processing was performed in cryoSPARC v4.3. Patch motion correction and Patch CTF estimation were run on all micrographs, followed by exposure curation with Manually Curate Exposures to remove low quality images, including micrographs with poor CTF fits, excessive motion, crystalline ice, empty or collapsed holes, extreme defocus or pronounced astigmatism, prominent support edges, or heavy ice or particulate contamination. An initial particle stack was extracted from a subset of one thousand curated micrographs using TOPAZ Extract with a pretrained model and subjected to one round of 2D classification. Selected classes were used to train a project specific TOPAZ model that was applied separately to the untilted and tilted

datasets. Because Pol γ particles are known to exhibit preferential orientation in cryo-EM, a complementary dataset at thirty degrees tilt was collected and processed in parallel. Particles from each set were extracted with a box size of 440 pixels with Fourier cropping to 110 pixels, classified in 2D, and used for ab initio reconstruction with three classes followed by heterogeneous refinement. For the untilted data one class was retained and for the tilted data two classes were retained. These particles were reextracted without Fourier cropping and processed through ab initio reconstruction with a single class and non-uniform refinement. Reference based motion correction was then performed and particles from both datasets were merged for a final non-uniform refinement that yielded a 2.64 Å map. The full processing workflow is summarised in Figure 37.

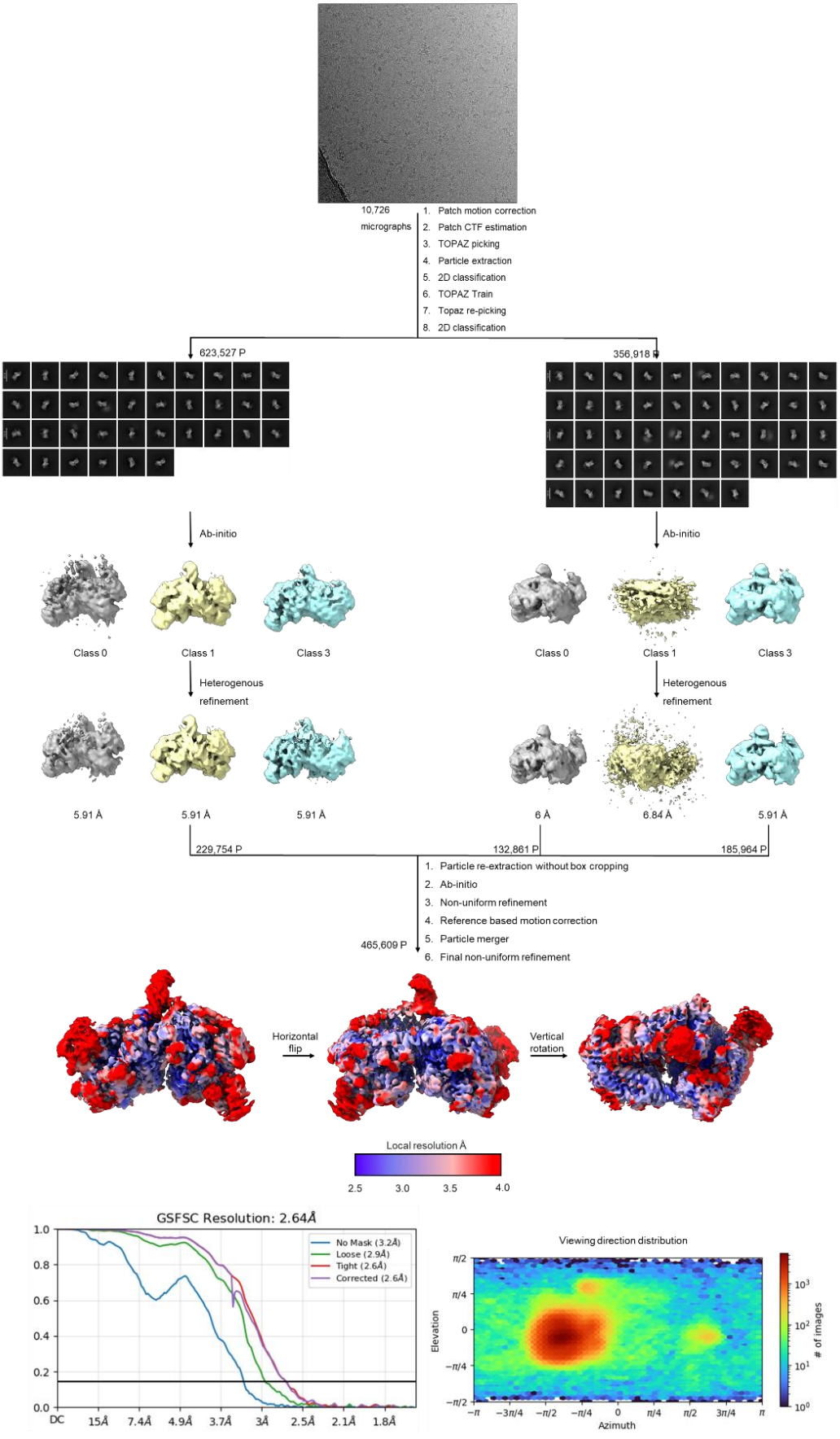

10,726 micrographs
1. Patch motion correction
2. Patch CTF estimation
3. TOPAZ picking
4. Particle extraction
5. 2D classification
6. TOPAZ Train
7. Topaz re-picking
8. 2D classification
623,527 P
356,918 P
Ab-initio
Ab-initio
Class 0
Class 1
Class 3
Class 0
Class 1
Class 3
Heterogenous refinement
Heterogenous refinement
5.91 Å
5.91 Å
5.91 Å
6 Å
6.84 Å
5.91 Å
229,754 P
132,861 P
185,964 P
1. Particle re-extraction without box cropping
2. Ab-initio
3. Non-uniform refinement
4. Reference based motion correction
5. Particle merger
6. Final non-uniform refinement
465,609 P
Horizontal flip
Vertical rotation
Local resolution Å
2.5
3.0
3.5
4.0
GSFSC Resolution: 2.64Å
No Mask (3.2Å)
Loose (2.9Å)
Tight (2.6Å)
Corrected (2.6Å)
1.0
0.8
0.6
0.4
0.2
0.0
DC
15Å
7.4Å
4.9Å
3.7Å
3Å
2.5Å
2.1Å
1.8Å
Viewing direction distribution
Elevation
π/2
π/4
0
−π/4
−π/2
−π
−3π/4
−π/2
−π/4
0
π/4
π/2
3π/4
π
Azimuth
# of images
10^3
10^2
10^1
10^0

Figure 37. Cryo-EM processing pipeline and final reconstruction for complex 1. Micrographs were motion and CTF corrected then curated, particles were picked with Topaz and processed by 2D classification, ab initio and heterogeneous refinement for the untilted and thirty degree tilted datasets. Particles were re extracted without Fourier cropping and merged after reference based motion correction for non-uniform refinement, yielding a 2.64 Å map. Lower panels show local resolution, GS FSC at the 0.143 criterion with a global resolution of 2.64 Å, and the viewing direction distribution indicating broad angular coverage.

Quality assessment of the final reconstruction is shown by the GS FSC and the viewing direction distribution. The GS FSC curves for No Mask, Loose, Tight, and Corrected meet at 2.64 Å at the 0.143 criterion, and the Tight and Corrected curves overlap, which indicates proper mask correction and rules out resolution inflation from masking. The viewing direction distribution shows broad and continuous angular coverage with one dominant cluster but no missing wedge. Because Pol γ particles are prone to preferred orientations, a thirty degree tilt dataset was included to broaden sampling. Although a residual orientation bias cannot be excluded, the overall map quality supports confident interpretation of the resolved features. Focused masking and 3D classification were applied to the region encompassing the RNA/DNA hybrid downstream of the gap, corresponding to the expected RNase H1 footprint. Neither masked classification nor local refinement recovered interpretable density for RNase H1. The map consistently showed well defined density for Pol γA engaging the hybrid, whereas features adjacent to the hybrid diminished with progressive processing and resolved as diffuse signal consistent with substrate mobility rather than a bound RNase H1. Occasional additional density near the hybrid observed in intermediate reconstructions was not retained upon refinement and is not supported by the final map.

Complex 2

In Complex 1 we reasoned that RNase H1 might require additional space around the hybrid to dock productively with Pol γ on the same substrate. The design therefore used a 5 nt gap and a downstream 10RNA/10DNA segment, conditions under which BLI confirmed that RNase H1 can bind even when the gap is as short as 2 nt (Figure 28). However, equilibrium binding does not guarantee a cryo-EM competent pose, and the final reconstruction lacked interpretable density for RNase H1, with features near the hybrid resolving as diffuse signal consistent with mobility. Because Pol γ is stably engaged with the hybrid, we considered that a short gap restricts the space and registers available for RNase H1, while the long downstream tract increases local flexibility, together promoting multiple transient RNase H1 orientations that average out during reconstruction. To reduce this heterogeneity while keeping both proteins on the substrate, we

moved to a geometry that gives RNase H1 more positioning space without disturbing the stable engagement of Pol γ.

Complex 2 was designed to test whether an 8 nt gap alleviates potential steric and geometrical constraints imposed by Pol γ at shorter gaps, thereby improving RNase H1 accommodation on the same substrate. We retained a 10 nt downstream RNA segment to limit hybrid mobility relative to longer RNA tracts, aiming to facilitate particle alignment and capture in cryo-EM. Guided by BLI observations that Pol γ and RNase H1 can co bind across gap lengths, we thought that an enlarged gap would favour a cryo-EM suitable binding geometry even if equilibrium binding is already supported at 2 nt. To stabilise a pre cleavage state without abolishing binding, I used wild-type RNase H1 in the presence of calcium, which inhibits RNase H1 activity while preserving complex formation (Rosta et al., 2014). The complex contained 2 µM Pol γ, 4 µM RNase H1, 2 µM RNA/DNA substrate (S7, Table 1) , 5 mM $CaCl_2$, and 1 mM dATP. Samples were vitrified and data were collected as described in Materials and Methods. Data were processed in cryoSPARC v4.5 (Figure 38).

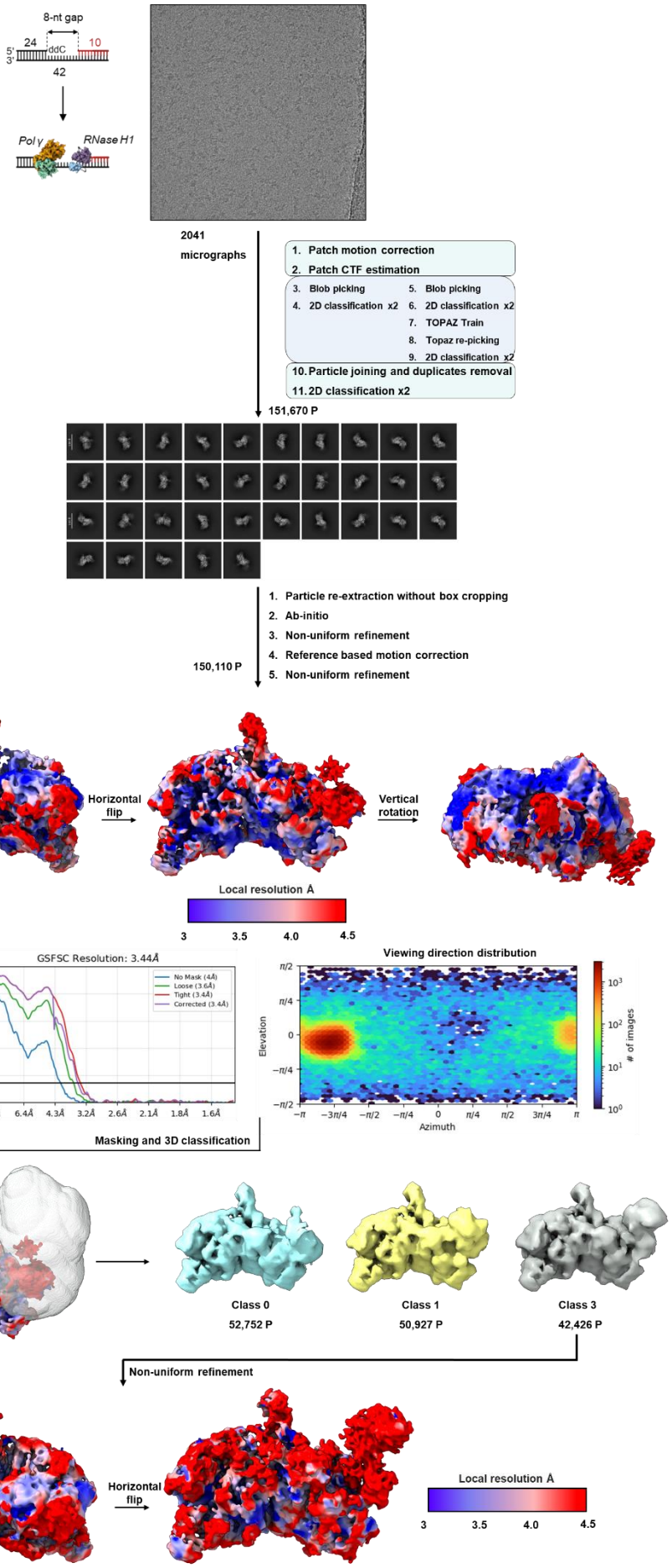

8-nt gap
24
ddC
10
5'
3'
42
Pol γ
RNase H1
2041
micrographs
1. Patch motion correction
2. Patch CTF estimation
3. Blob picking
4. 2D classification x2
5. Blob picking
6. 2D classification x2
7. TOPAZ Train
8. Topaz re-picking
9. 2D classification x2
10. Particle joining and duplicates removal
11. 2D classification x2
151,670 P
1. Particle re-extraction without box cropping
2. Ab-initio
3. Non-uniform refinement
4. Reference based motion correction
5. Non-uniform refinement
150,110 P
Horizontal flip
Vertical rotation
Local resolution Å
3
3.5
4.0
4.5
GSFSC Resolution: 3.44Å
No Mask (4Å)
Loose (3.6Å)
Tight (3.4Å)
Corrected (3.4Å)
Viewing direction distribution
Elevation
Azimuth
# of images
Masking and 3D classification
Class 0
52,752 P
Class 1
50,927 P
Class 3
42,426 P
Non-uniform refinement
Horizontal flip
Local resolution Å
3
3.5
4.0
4.5

Figure 38. Cryo-EM processing pipeline and final reconstruction for complex 2. Micrographs were corrected for motion and CTF and curated, then particles were identified by Blob picker and a project trained Topaz model and subjected to ounds of 2D classification, ab initio reconstruction and non-uniform refinement. Particles were re extracted without Fourier cropping, refined with reference based motion correction and a second non-uniform refinement, yielding a reconstruction of the Pol γ core. Lower panels show local resolution, the gold standard FSC at the 0.143 criterion with a global resolution of 3.44 Å, and the viewing direction distribution that reveals pronounced orientation bias. Focused 3D classification over the expected RNase H1 footprint produced three classes, one of which displayed extra density next to the hybrid that is compatible with RNase H1 but lacks resolvable secondary structure.

Patch Motion Correction and Patch CTF estimation were performed on all micrographs, followed by manual exposure curation to remove low quality images. Initial particle picking was carried out using the Blob picker, and selected two dimensional classes from this dataset were used to train a project specific TOPAZ model. Particles extracted by TOPAZ were then refined through two additional rounds of two dimensional classification, after which the TOPAZ stack was merged with the Blob picker set and duplicates were removed. The final merged stack was re extracted without Fourier cropping in order to retain the full frequency content for downstream processing. Ab initio reconstruction with a single class produced a sensible volume, while attempts with multiple classes converged to nearly identical solutions, indicating a relatively homogeneous particle population at this resolution. This volume was subsequently refined by non-uniform refinement, followed by Reference based Motion Correction to recover high frequency information and a second non-uniform refinement to consolidate these corrections. The resulting map displayed the main body of Pol γ with improved separation of structural features, and the overall quality of the reconstruction was assessed using the gold standard FSC curves. The No Mask, Loose, Tight, and Corrected curves converged consistently at the 0.143 criterion, with Tight and Corrected overlapping, which indicates appropriate mask correction and rules out resolution artefacts from masking. The 2D classes revealed clear views of the Pol γ, and ab initio reconstruction consistently produced only one distinct solution, suggesting that the dataset captured a stable particle state at the available signal to noise ratio. To assess the contribution of RNase H1, focused 3D classification was applied with a mask placed over the expected RNase H1 footprint downstream of the substrate gap. Partitioning the dataset into three classes of approximately equal size yielded one class displaying an additional density adjacent to the hybrid, consistent in shape and position with an RNase H1 body. However, this volume did not resolve secondary structure features such as helices or beta sheets, which is expected given the moderate dataset size and the 200 kV collection conditions. Analysis of viewing directions showed a strong orientation preference with dominant bands in azimuth, and the

sampling compensation factor was 0.304, indicating substantial angular bias and under representation of certain orientations. This preferred orientation limits angular completeness and reduces the ability to resolve peripheral features, including RNase H1. Overall, the Complex 2 dataset suggests that extending the gap to 8 nt can sometimes reveal extra density next to the hybrid in focused classification, but the lack of clear secondary structure means we cannot say for certain that this corresponds to RNase H1. The limitations of this dataset arise from its modest size, the significant orientation bias, and the use of a screening level 200 kV microscope. These considerations led to the decision to acquire a larger dataset on a Titan Krios microscope to improve angular coverage and resolution, with the aim of resolving RNase H1 secondary structure elements. The results of this high quality dataset are presented in the following section. However, processing of the Titan Krios dataset did not yield interpretable density for RNase H1. Despite the larger data volume and improved separation of features within the Pol γ reconstruction, focused three dimensional classification over the expected RNase H1 footprint did not recover a stable body. The viewing direction distribution indicated persistent orientation preference, and peripheral densities adjacent to the hybrid diminished with progressive refinement, consistent with substrate mobility or reduced occupancy of RNase H1. Thus, while the polymerase is well resolved, the dataset does not support a unique RNase H1 pose under the tested conditions.

7.2.8. Structure prediction

Given that the cryo-EM datasets did not yield an averaged RNase H1 body, I used the AlphaFold Server implementation of AlphaFold3 to obtain a predictive arrangement of Pol γ and RNase H1 on the gapped RNA/DNA substrate used in Complex 2. To mirror the experimental condition I included calcium ions and dATP in the input. Predictions were run with default server settings for complex modelling. The server returned a model with the standard confidence outputs (Figure 39). This prediction is presented as a conservative hypothesis consistent with the functional data. It is not derived from cryo-EM density and does not assign a unique RNase H1 pose.

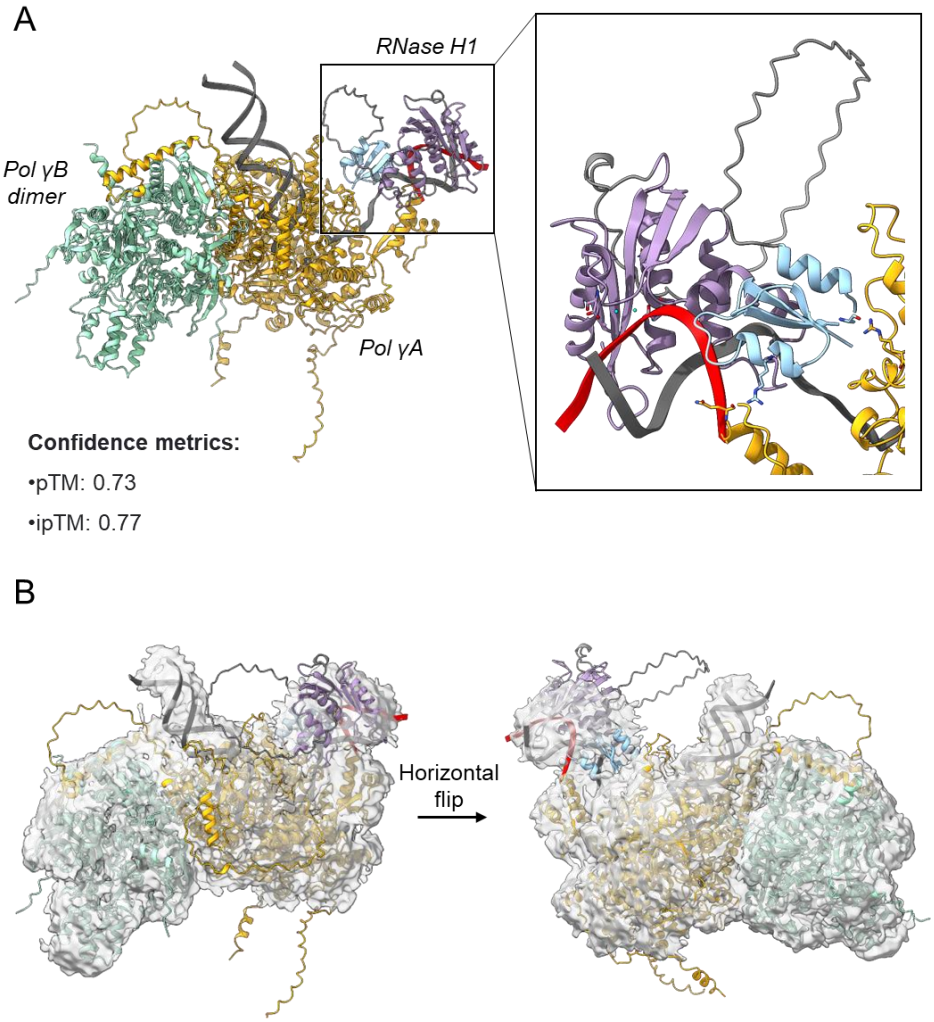


Figure 39. AlphafoldFold3 prediction of the Pol γ-RNase H1-RNA/DNA complex. (A) AlphaFold3 model of the ternary complex. PolγA is orange, the PolγB dimer is green, RNase H1 is purple with HBD in light blue and the linker in dark grey, the RNA/DNA hybrid is dark grey with the RNA segment in red. The boxed inset shows a close up of RNase H1 positioned on the hybrid. Confidence metrics reported by the AlphaFold Server are pTM 0.77 and ipTM 0.73. (B) The predicted model of the complex is placed as a rigid body into the cryo-EM density from Complex 2; the map is displayed as a translucent surface. The right view shows the same placement after a horizontal flip.

The AlphaFold Server reported ipTM equal to 0.73 and pTM equal to 0.77 for the predicted complex. Per residue confidence is high to very high across the cores of Pol γA, Pol γB, and the RNase H1 catalytic domain, and lower on flexible loops and on the nucleic acid, especially at termini. In the AlphaFold3 arrangement RNase H1 sits on the RNA/DNA hybrid with the HBD approaching the minor groove and the catalytic domain oriented toward the scissile phosphate,

as shown in Figure 28, panel A. The recognition patch on HBD that includes Y29, W43, F58, R57, and K60 lies along the DNA phosphate ribose contour, while the RNA binding loop from D51 to A56 follows the row of ribose 2' OH groups on the RNA strand, which is consistent with published structures of human RNase H1 (Nowotny, 2008). The catalytic site positions the scissile phosphate beside the conserved acidic cluster D145, E186, D210, and D274 with H264 on a mobile loop near the product site, and the geometry is compatible with the two metal ion mechanism (Nowotny 2007), in which the acidic side chains coordinate the metals and the metals bridge to the phosphate and the water nucleophile. Only HBD mediated contacts between RNase H1 and PolγA are observed, occurring in a sterically crowded region with an explicit atomic clash in which part of PolγA overlaps the nucleic acid, therefore this interface is considered uncertain and likely a placement artefact characteristic of AF3 predictions for large assemblies. Regarding Pol γ, the AlphaFold3 model positions the DNA primer template segment within the canonical DNA binding cleft of PolγA, aligning the 3′ primer terminus with the active site, with the PolγB dimer arranged alongside the surface of PolγA, which is in line with known holoenzyme architecture. Inspection of the cryo-EM map from Complex 2 shows a contiguous density adjacent to the position predicted for RNase H1, as illustrated in Figure 39, panel B. This observation indicates geometric compatibility between the prediction and the experimental map but does not establish the presence or orientation of RNase H1 in that reconstruction and should be regarded as supportive rather than conclusive.

7.2.9. Summary

Figure 25 document that the recombinant proteins used in this study were purified and verified to be correctly folded, establishing suitable inputs for the experiments that follow. Surface plasmon resonance then defined a direct interaction between RNase H1 and both Pol γ subunits, with the catalytic domain of RNase H1 mediating binding in the submicromolar range, while the hybrid-binding domain bound much more weakly. In particular, the catalytic domain bound Pol γA and Pol γB with $K_d \approx 0.29$ μM and $K_d \approx 0.63$ μM, whereas the hybrid-binding domain showed $K_d \approx 29$ μM and $K_d \approx 139$ μM, respectively, indicating that the catalytic domain may be the principal interface for complex formation (Figure 26). Functional assays on a gapped RNA/DNA substrate showed that Pol γ alone completed a 10 nt gap within 10 min at 25 °C, while catalytically inactive RNase H1 D210N caused a robust pause after incorporation of 7 nt, with accumulation of a 31 nt intermediate. The D210N variant of the isolated catalytic domain produced only a transient pause and full gap-filling was recovered, implicating the hybrid-binding domain in prolonged retention of inactive RNase H1 on the hybrid and in the resulting polymerase stall (Figure 27). Biolayer interferometry confirmed that Pol γ and RNase H1-D210N can stably co bind the same substrate across 2, 3, and 8 nt gaps and that the order of

protein addition does not affect association or dissociation, indicating that the stall does not arise from Pol γ dissociation (Figures 28 to 30). Coupled gap-filling-nuclease assays showed that Pol γ enhances RNase H1 cleavage to fully remove RNA from the downstream hybrid, an effect that scales with RNase H1 concentration, is not reproduced by Pol γB alone, and is strongly reduced when dNTPs are omitted or when primer extension is blocked with ddC, consistent with facilitation by gap shortening. The catalytic domain of RNase H1 is also stimulated by Pol γ, but with lower efficiency than the full-length enzyme, and activity depends on RNA tract length, with no cleavage detected for a two-ribonucleotide tract (Figures 31 to 34). Size exclusion chromatography mixture of Pol γ, RNase H1, and substrate demonstrated a clear elution shift and co-elution of all the components, consistent with a stable ternary assembly (Figure 35). cryo-EM microscopy yielded a 2.64 Å map of Pol γ with broad angular coverage after combining untilted and thirty degree tilted data for Complex 1, yet no reproducible density corresponding to RNase H1 was observed. In Complex 2, collected at 200 kV, an additional volume adjacent to the hybrid was sometimes seen but lacked secondary structure features, and orientation bias remained substantial, limiting interpretability (Figures 37 and 38). An AlphaFold 3 model of the Complex 2 arrangement returned pTM 0.77 and ipTM 0.73 and placed RNase H1 on the hybrid with a pose that is geometrically compatible with a contiguous density in the map (Figure 39). Together, these findings support a mechanism where direct interaction and polymerase-dependent synthesis promote RNase H1 mediated primer removal, while prolonged retention of catalytically inactive RNase H1 impedes Pol γ. My efforts to obtain high-resolution structural data on Pol γ-RNaseH1-RNA/DNA complex architecture have so far been unsuccessful due to the experimental limitations in visualisation of a unique RNase H1 pose on the substrate.

8. Discussion

8.1. Part I: Functional cooperation of RNase H1 and EXOG in mitochondrial RNA primer removal

The first part of this thesis sought to elucidate a critical, yet incompletely understood, step in mtDNA replication: the removal of the terminal ribonucleotide remaining after initial processing by RNase H1. Being vital for creating an RNA-free nick for DNA Ligase III, this mechanism was hypothesised to be executed not by stochastic enzymatic activity but by a coordinated pathway involving the mitochondrial exonuclease EXOG pathway (Al-Behadili et al., 2018; Wu et al., 2019; Karlowicz et al., 2022). My experimental approach, integrating rigorous biochemical quantification, interaction mapping, functional reconstitution, and structural prediction, has contributed to the understanding of this essential two-nuclease module

Defining the functional coordination

Using SPR, I confirmed a direct, high-affinity interaction between full-length RNase H1 and immobilised EXOG, characterised by K_d of 0.38 ± 0.13 μM. This stable physical association supports the formation of a dedicated "primer processing module". Further evidence came from glycerol gradient co-sedimentation, where the RNase H1 co-migrated with the heavier EXOG dimer, confirming complex stability in solution. This physical coupling appears to influence the functional outcome. The two-nuclease cleavage assay indicated a division of activities, where RNase H1 performs the initial partial cleavage and EXOG provides the final processing. EXOG alone showed minimal activity on the longer RNA/DNA hybrid substrate, consistent with its reliance on RNase H1 to produce a suitable shortened intermediate. This synergistic model likely promotes efficient channeling of the partially processed intermediate, thereby reducing premature release and facilitating removal of the terminal RNA dinucleotide. Collectively, these observations support a potential flap-independent pathway for primer maturation, in contrast to flap-dependent models previously proposed for mitochondrial nucleases such as MGME1 and FEN1.

Interpreting pathogenic mutants and compensation

Aim 1 also required assessing how disease-linked variants perturb this coordination. My analysis of the RNase H1 pathogenic mutants V142I and A185V, both linked to adult-onset mitochondrial encephalomyopathy (Reyes et al., 2015), indicated defects associated with structural stability. Initial biophysical characterisation showed marked structural destabilisation in both variants, particularly A185V, which exhibited a substantially reduced melting temperature (T_m = 32.1 °C).

In contrast to previous reports of impaired function, the V142I variant displayed catalytic activity in my cleavage assay that was indistinguishable from the wild-type enzyme. This discrepancy can likely be explained by the different experimental conditions. My assay was performed over a short duration (5 minutes) at 25 °C, whereas studies reporting reduced activity employed longer incubations (60 minutes) at 37 °C (Al-Behadili et al., 2018). Therefore, my results suggest that the V142I variant retains immediate catalytic competence under the tested conditions but is structurally fragile. The brief, sub-physiological temperature conditions of my assay likely capture its function before significant thermal denaturation occurs, whereas prolonged exposure to 37 °C in other studies reveals its functional collapse over time. This distinction is important, as the pathogenicity of V142I *in vivo* may result from a progressive loss of structural stability rather than from an inherently deficient active site. This aligns with the hypothesis that its pathogenic effect may stem from dysregulated behaviour rather than a simple loss of function (Holt, 2019).

Elucidating the molecular interfaces

Aim 2 focused on identifying and characterising the molecular interfaces governing the RNase H1-EXOG interaction through an integrated computational-biochemical approach. To delineate the interaction interface, I performed SPR analysis using a series of truncated RNase H1 variants. The results implicate the catalytic domain (CTD) as the main mediator of the interaction. The CTD alone retained substantial binding affinity for EXOG (K_d = 0.77 ± 0.31 μM), whereas the hybrid-binding domain (HBD) was largely insufficient for binding (K_d = 12.6 ± 10.1 μM). The full-length enzyme bound most tightly (K_d = 0.38 ± 0.13 μM), while inclusion of the flexible linker between the CTD and HBD partially restored affinity. These results support a model in which the CTD likely serves as the main docking interface, positioning the catalytic sites in proximity to facilitate substrate transfer. Furthermore, my kinetic analysis indicated that the flexible linker modulates binding affinity mainly by increasing the association rate (k_{on}), suggesting it facilitates the initial recognition rather than direct binding.

Computational insights and interface redundancy

Technical limitations in crystallisation motivated the use of AlphaFold2 Multimer to model the RNase H1-EXOG interaction. The resulting models were consistent with the biochemical data: the CTD-only complex yielded the highest interface confidence (ipTM = 0.650), whereas confidence was lower for full-length RNase H1 (ipTM = 0.553), likely reflecting its intristic flexibility. These models guided targeted mutagenesis, highlighting conserved residues predicted to stabilase the complex. However, subsequent experimental validation using single alanine substitutions failed to disrupt the interaction. This result is informative, suggesting that the

interface may represent a broad, structurally redundant network rather than relying on a few critical residues. Such redundancy may be a feature of essential biological systems, providing robustness against the loss of individual interactions.

Challenges in structural determination

Aim 3, focused on determining the three-dimensional structures of the key complexes, proved challenging but yielded valuable insights into the dynamics of RNase H1. Despite extensive crystallisation trials, a high-resolution structure of the RNase H1-EXOG complex could not be obtained. Molecular replacement consistently identified only the rigid EXOG dimer, with no interpretable electron density corresponding to RNase H1. This likely reflects the protein's high conformational flexibility, which prevents ordered crystal packing. This limitation nevertheless provides structural context for the functional observations. Eukaryotic RNase H1 differs from its bacterial counterpart partly due to the evolutionarily acquired HBD, linked by a flexible linker (Cerritelli & Crouch, 2009). This inherent mobility has long made full-length RNase H1 a difficult subject for structural biology, with no complete structures available to date (Hyjek et al., 2019). Previous studies have noted the considerable difficulty in crystallising even its individual domains (Nowotny et al., 2008), and early attempts to solve the structure of the full-length protein were unsuccessful, necessitating the use of isolated catalytic domains for structural work (Nowotny et al., 2007). While detrimental to static structural determination, this flexibility is likely essential for RNase H1 functional versatility, enabling it to dynamically position its catalytic domain to interact with partners such as EXOG and Pol γ. The structural challenge is consistent with a mechanistic model in which the RNase H1-EXOG module is functionally dynamic rather than static. This reflects a broader challenge in the field, as the dynamic and often transient nature of interactions between enzymes involved in mtDNA maintenance has complicated efforts to establish a single, definitive model of mtDNA replication (Yasukawa & Kang, 2018). Future high-resolution analyses will likely benefit from methods optimised for flexibility, such as cryo-EM combined with covalent cross-linking.

8.2. Part II: Pol γ-mediated regulation of RNase H1 in RNA primer processing

Building on previous work that identified a direct interaction between RNase H1 and Pol γ (Karlowicz et al 2022), the second part of this thesis aimed to dissect the molecular basis and functional consequences of this interaction. Through an integrated approach combining biophysical, biochemical, and structural methods, I sought to define the physical interface, uncover the functional interdependence of the two enzymes, and resolve the structure of the complex.

Defining the physical interaction between RNase H1 and Pol γ

To better understand the functional coordination between RNase H1 and Pol γ, I sought to validate and map their physical interaction. I employed SPR to quantitatively characterise the binding between RNase H1 constructs and the individual subunits of the Pol γ holoenzyme. These experiments demonstrated a direct and stable interaction, with full-length RNase H1 binding to both the catalytic subunit Pol γA (K_d = 0.15 ± 0.03 μM) and the accessory subunit Pol γB (K_d = 0.36 ± 0.08 μM). Domain mapping indicated that this interaction is primarily mediated by the C-terminal catalytic domain of RNase H1. The isolated CTD bound to Pol γA and Pol γB with affinities (K_d = 0.29 ± 0.07 μM and K_d = 0.63 ± 0.04 μM, respectively) comparable to the full-length enzyme, whereas the N-terminal hybrid-binding domain (HBD) exhibited only a very weak binding. This finding is notable, as it places the protein-protein interface within the catalytic core of RNase H1 rather than its substrate-binding module. This architecture is consistent with a model in which the interaction coordinates the catalytic activities of the two enzymes in proximity on a shared nucleic acid substrate.

Uncovering the functional interdependence of RNase H1 and Pol γ

Having established a direct physical link, I next examined the functional consequences of this interaction using reconstituted enzymatic assays. My experiments revealed a reciprocal relationship in which each enzyme quantitatively modulates the activity of the other. To simulate defective RNA primer processing, I tested the effect of a catalytically inactive RNase H1 mutant (D210N) on Pol γ gap-filling activity. In the presence of RNase H1-D210N, Pol γ activity was markedly reduced, stalling after incorporation of seven of ten nucleotides. These observations are consistent with the inactive RNase H1 acting as a physical roadblock on the DNA template. Further analysis showed that this stalling depends on the HBD: when the isolated CTD-D210N was used, the block was transient. This implicates HBD, with its high affinity for the RNA/DNA hybrid, in stabilising the inactive enzyme on the substrate and forming a persistent barrier. BLI experiments supported this model, showing that both enzymes can stably co-occupy the same template. These findings have pathological implications, suggesting that catalytically impaired but substrate-binding RNase H1 variants may not only fail to remove RNA but could also impede replication fork progression.

Pol γ stimulates complete RNA removal by RNase H1

Conversely, I investigated whether Pol γ activity could influence the function of wild-type RNase H1. Although RNase H1 alone efficiently cleaves an RNA primer, it typically leaves one to three ribonucleotides at the RNA-DNA junction. The addition of active Pol γ and dNTPs led to

accumulation of a fully cleaved, RNA-free product, indicating that Pol γ stimulates complete primer removal. This stimulation was dependent on the polymerase activity of Pol γ, as the effect was absent when Pol γB was used alone, when dNTPs were omitted, or when primer extension was blocked with ddC. These findings suggest a mechanism in which gap-shortening by Pol γ facilitates removal of the terminal ribonucleotides by RNase H1. This coordinated action suggests an efficient mechanism that ensures generation of a ligatable DNA nick. Even the isolated CTD of RNase H1 was stimulated by Pol γ, albeit less efficiently, suggesting that, while the CTD serves as the principal interface for protein-protein interaction, the HBD is essential for optimal activity.

This observation raises an important question: as the Pol γ-RNase H1 partnership is proficient in complete RNA removal, why was there no activity on a 2-nt RNA substrate and reduced activity on a 4-nt RNA substrate? The most plausible explanation lies in the requirements for establishing a cooperative complex. The ability of RNase H1 to excise the final ribonucleotides appears to depend on a cooperative state that arises only within a dynamic, processive replication complex. On longer substrates, the HBD domain of RNase H1 can bind with high affinity, allowing Pol γ to engage. Within this moving complex, Pol γ may promote a conformational change in RNase H1 that could facilitate cleavage of the final ribonucleotides. However, a 2-4 nt RNA hybrid is too short to provide an effective binding platform for RNase H1. Without that stable engagement, the synergistic interaction with Pol γ does not occur, and RNase H1 defaults to its intrinsic inability to process such minimal substrates. Therefore, while my results define an efficient pathway for primer removal driven by the replisome, they also delineate its limits. This offers a mechanistic rationale for the involvement of a secondary nuclease such as EXOG, which could resolve primers left behind if the primary Pol γ-RNase H1 complex dissociates prematurely.

Structural analysis of the Pol γ-RNase H1 Complex

A primary objective was to visualise this dynamic interplay by determining the structure of the Pol γ-RNase H1-RNA/DNA hybrid ternary complex. Size-exclusion chromatography confirmed that the three components form a stable assembly in solution, validating it as a target for structural studies. Despite this biochemical stability, extensive cryo-EM analyses did not yield a resolved structure of the complete complex. I consistently obtained high-resolution reconstructions of the Pol γ holoenzyme bound to the DNA substrate (up to 2.64 Å) but failed to resolve interpretable density for RNase H1. The diffuse density observed at its expected position is consistent with high mobility. Although no ordered density was obtained, this result provides mechanistic insight suggesting that the interaction between Pol γ and RNase H1 is inherently

flexible. This flexibility, likely conferred by the long linker connecting the HBD and CTD of RNase H1, is probably essential for its function, enabling the enzyme to scan the substrate and dynamically engage with Pol γ at the replication fork.

To address this, I attempted to form the complex using the isolated CTD. However, no interpretable density was obtained, likely because its small molecular mass (16.9 kDa) limits the alignment signal relative to the large Pol γ holoenzyme. While this limitation identifies the technical barriers to be addressed in future work, a predicted model of the assembly was generated using AlphaFold3. The resulting model positions RNase H1 on the RNA/DNA hybrid in a catalytically competent orientation, forming plausible, yet low-confidence, contacts with Pol γ. This predicted interface warrants comparison with experimentally determined interaction sites for other replisome components. A recent cryo-EM study (Riccio et al., 2024) successfully resolved the structure of Pol γ in complex with the mitochondrial single-stranded DNA binding protein (mtSSB). In their reconstruction, they identified a distinct density proximal to the catalytic subunit, Pol γA, which they assigned to mtSSB, thereby mapping a key protein-protein interface within the replisome to residues 993-1058 on Pol γA. This region is noted to be of critical importance, containing several disease-causing missense mutations that suggest its functional significance. It is plausible that the flexible interaction between RNase H1 and Pol γ, which proved challenging to resolve in my own cryo-EM analysis, occurs at this same functionally critical interface. The low-confidence contacts suggested by the AlphaFold3 model could be tested against this experimentally defined mtSSB binding site. If this were the case, it would imply that RNase H1 and mtSSB may compete for binding to Pol γ, or that their interactions are temporally regulated during different stages of primer processing and lagging-strand synthesis. This competition or coordinated exchange could be a crucial regulatory mechanism at the replication fork, a hypothesis that provides a clear direction for future mutational and structural studies

8.3. General Discussion

The findings presented in this thesis delineate two distinct, yet functionally interconnected, modules for mitochondrial RNA primer removal, both centred on RNase H1. The RNase H1-EXOG partnership functions as a specialised, sequential processing unit, where RNase H1 performs bulk degradation and EXOG executes the final excision. In contrast, the RNase H1-Pol γ partnership represents a more dynamic and integrated regulatory module within the core replisome.

A key finding of this work is that the Pol γ-RNase H1 complex can, under specific conditions, achieve complete primer removal. This observation initially appeared suprising, as the complex was inactive on a short 2-nt substrate mimicking the final ligation intermediate. This apparent discrepancy can be reconciled by a mechanistic insight from Part II: the capacity for complete cleavage is not an intrinsic property of RNase H1 but an emergent one of the processive Pol γ-RNase H1 complex. The dynamic, gap-shortening action of Pol γ appears to modulate RNase H1 activity, but only during active translocation along a sufficiently long substrate. A short, static substrate fails to provide the necessary "runway" for this cooperative state to be established.

This model provides a coherent framework for understanding the system's redundancy. It suggests that the RNase H1-EXOG module functions not merely as a backup but as a parallel, context-dependent pathway for intermediates that the primary processive machinery is not optimised to engage de novo. The replisome, geared for speed, may not achieve complete primer removal in every cycle. The residual 1-3 nt fragments would therefore not represent failure but rather reflect a system prioritising processivity. These static intermediates then become the preferred substrates for the RNase H1-EXOG complex. Such a dual strategy could combine speed with accuracy, providing resilience to mitochondrial DNA replication.

An emerging aspect from this work is the characterisation of RNase H1 as a dynamically regulated enzyme whose activity is strongly modulated by its protein partners. This model of partner-dependent regulation is supported by established literature on single-stranded DNA-binding proteins such as RPA, which can transform RNase H1 into a more processive exoribonuclease (Li et al., 2024). The stimulatory effect of Pol γ observed here is mechanistically analogous, framing Pol γ as a dual-function component of the replisome, both a polymerase and a regulatory factor for its nuclease partner. Furthermore, the architectural role of mtSSB is likely important for coordinating this process *in vivo* by stabilising R-loops and protecting displaced DNA, thereby presenting the nucleic-acid structures upon which the Pol γ-RNase H1 complex acts. My findings extend this concept to extrinsic regulation, where the binding of Pol γ and EXOG to the RNase H1 catalytic domain may promote specific functional states of the enzyme.

Although the inherent flexibility of RNase H1, evident from unsuccessful high-resolution structural analyses with two different partners, presented a major challenge, this integrated model yields several testable predictions. The functional flexibility of RNase H1 could be captured structurally by employing covalent cross-linking strategies to trap these dynamic complexes for cryo-EM analysis. Finally, the context-dependent engagement of these pathways *in vivo* could be explored using proximity-ligation assays to compare the relative association of RNase H1 with Pol γ versus EXOG under different cellular conditions.

In conclusion, this thesis refines our understanding of mitochondrial primer removal, moving beyond a simple linear pathway to reveal a coordinated network of redundant, specialised, and dynamically regulated enzymatic modules that act together to safeguard the integrity of the mitochondrial genome.

9. Scientific meetings and training

2025

- **Piotr Purzycki**, Anna Karłowicz, Michał R Szymański. *Structural basis for functional cooperation between Pol γ and RNase H1 proteins in the processing of mitochondrial RNA primers*. 6th BIO Life Science Congress. 17-20 September 2025, Poznań, Poland.
- Beata M. Walter, Adela Bledea, **Piotr Purzycki**, Paulina Nastały, Katarzyna Bury, and Michał R. Szymanski. *Lipid-Mediated Regulation of Human Mitochondrial Ligase IIIα by Cardiolipin and PolγA*. 6th BIO Life Science Congress. 17-20 September 2025, Poznań, Poland.

2024

- **Piotr Purzycki**, Michał R Szymański. *Structural basis for functional cooperation between proteins involved in the processing of mitochondrial RNA primers*. The 7th DNA Polymerases Meeting. 28-31 August 2024, Warsaw, Poland. Poster presentation.
- Julia Karasińska, Katarzyna A. Paniak, **Piotr Purzycki**, Anna Karłowicz, Andrzej B Dubiel, Jolanta Czerwińska, Bartosz Szczęsny, Roman J Szczęsny, Michał R Szymański. *Investigating APE1-EXOG synergy in mitochondrial DNA repair.* EMBO Workshop: Molecular Biology of Mitochondrial Gene Maintenance and Expression. 19-23 May 2024, Józefów, Poland.

2023

- XXVI Biotechnology Summer School: RNA in Biology and Medicine. 11-15 September 2023, Kościerzyna, Poland. Participant.

2022

- **Piotr Purzycki**, Anna Karłowicz, Andrzej Dubiel, Adela Bladea, Michał R Sztmański. *Cooperation of Polγ and RNase H1 in the processing of mitochondrial RNA primers*. EMBO Practical Course: Characterization of Macromolecular Complexes by Integrative Structural Biology. 28 May-4 June 2022, Grenoble, France.
- Anna Karlowicz, Andrzej B Dubiel, Adela Bledea, **Piotr Purzycki**, Marta Grzelewska, Ryan J McAuley, Bartosz Szczesny, Jolanta Czerwinska, Roman J Szczesny, Michal R Szymanski. *In vitro reconstitution reveals a key role of human mitochondrial EXOG in RNA primer processing.* EMBO Workshop: Molecular Biology of Mitochondrial Gene Maintenance and Expression. 15-19 May 2022, Bro, Sweden.

- Andrzej B Dubiel, Anna Krałowicz, Ryan J McAuley, Adela Bledea, **Piotr Purzycki**, Bartosz Szczesny, Michal R Szymanski. *Identification and structural analysis of newly discovered mitochondrial protein complexes.* EMBO Workshop: Molecular Biology of Mitochondrial Gene Maintenance and Expression. 15-19 May 2022, Bro, Sweden.
- **Piotr Purzycki**, Anna Karłowicz, Andrzej Dubiel, Adela Bladea, Michał R Szymański. *Structural basis for functional cooperation between Pol γ and RNase H1 proteins in the processing of mitochondrial RNA primers*. iNEXT-Discovery 2nd Annual Scientific Meeting. 29-30 August 2022, Warsaw, Poland.
- Andrzej B Dubiel, Anna Krałowicz, Ryan J McAuley, Adela Bledea, **Piotr Purzycki**, Bartosz Szczesny, Michal R Szymanski. *Identification and structural analysis of newly discovered mitochondrial protein complexes*. iNEXT-Discovery 2nd Annual Scientific Meeting. 29-30 August 2022, Warsaw, Poland.
- Anna Karlowicz, Andrzej B Dubiel, Adela Bledea, **Piotr Purzycki**, Marta Grzelewska, Ryan J McAuley, Bartosz Szczesny, Jolanta Czerwinska, Roman J Szczesny, Michal R Szymanski. *In vitro reconstitution reveals a key role of human mitochondrial EXOG in RNA primer processing*. iNEXT-Discovery 2nd Annual Scientific Meeting. 29-30 August 2022, Warsaw, Poland.
- Beata Walter, **Piotr Purzycki**, Michał R Szymański. *DNA ligase III interactions in human mitochondrial DNA integrity and stability*. Helmholtz Training Workshop on Integrative Structural Biology. 8-14 October 2022, Berlin-Adlershof & Berlin-Buch, Germany.

2021

- **Piotr Purzycki**, Anna Karłowicz, Andrzej Dubiel, Adela Bladea, Michał R. Szymański. *Structural basis for functional cooperation between proteins involved in the processing of mitochondrial RNA primers*. EMBO Young Scientists' Forum. 21-22 October 2021, Warsaw, Poland.
- Anna Karłowicz, **Piotr Purzycki**, Michał R Szymański. *EXOG-mediated RNA cleavage is a limiting step for complete primer removal in human mitochondria*. EMBO Young Scientists' Forum. 21-22 October 2021, Warsaw, Poland. Co-author of the poster presentation.
- Andrzej B Dubiel, **Piotr Purzycki**, Michał R Szymański. *Application of in vitro techniques to search for proteins which could be involved in mitochondrial DNA metabolism*. EMBO Young Scientists' Forum. 21-22 October 2021, Warsaw, Poland. Co-author of the poster presentation.

Outreach and public engagement

- Open Days and Science Picnic, Intercollegiate Faculty of Biotechnology University of Gdańsk and Medical University of Gdańsk. Volunteer involvement in promotional and educational activities for prospective students and the general public. 2022.

10. List of Publications

Peer-reviwed Articles:

- Katarzyna Wegrzyn, Monika Oliwa, Marzena Nowacka, Elżbieta Zabrocka, Katarzyna Bury, **Piotr Purzycki**, Paulina Czaplewska, Justyna Pipka, Rafael Giraldo, Igor Konieczny, Rep protein accommodates together dsDNA and ssDNA which enables a loop-back mechanism to plasmid DNA replication initiation, Nucleic Acids Research, Volume 51, Issue 19, 27 October 2023, Pages 10551–10567, https://doi.org/10.1093/nar/gkad740
- Anna Karlowicz, Andrzej B Dubiel, Jolanta Czerwinska, Adela Bledea, **Piotr Purzycki**, Marta Grzelewska, Ryan J McAuley, Roman J Szczesny, Gabriela Brzuska, Ewelina Krol, Bartosz Szczesny, Michal R Szymanski, In vitro reconstitution reveals a key role of human mitochondrial EXOG in RNA primer processing, Nucleic Acids Research, Volume 50, Issue 14, 12 August 2022, Pages 7991–8007, https://doi.org/10.1093/nar/gkac581

Manuscripts in preparation

- **Piotr Purzycki**, Anna Karłowicz, Michal R. Szymanski. DNA Polymerase γ enhances RNase H1 activity in a reconstituted human mitochondrial replication system. Manuscript in preparation.
- Marta Wyszkowska, Andrzej B. Dubiel, **Piotr Purzycki**, Douglas Houston, Bartosz Szczęsny, Michal R. Szymanski. Inhibition of mitochondrial nuclease EXOG to improve chemotherapy. Manuscript in preparation.
- Beata M. Walter, Adela Bledea, **Piotr Purzycki**, Paulina Nastały, Katarzyna Bury, and Michał R. Szymanski. Lipid-Mediated Regulation of Human Mitochondrial Ligase IIIα by Cardiolipin and PolγA. Manuscript in preparation.

11. Technical Skills and Responsibilities Developed During My PhD

During my doctoral training I had the opportunity to work with a wide range of experimental techniques and laboratory equipment that extended beyond the methodologies presented in this thesis. In addition to the approaches described in the main chapters, I gained experience with the following techniques and laboratory responsibilities

- fluorescence anisotropy assays for analysing protein-protein and protein nucleic acid interactions
- work with radioactive isotopes
- dynamic light scattering (DLS)
- chemical crosslinking of protein-protein and protein-RNA complexes
- cryo-EM sample preparation, grid freezing, grid screening, data collection and processing
- operation and maintenance of fast protein liquid chromatography systems (FPLC), including NGC (Biorad) and AKTA (Cytiva)
- urea-PAGE, SDS-PAGE, native-PAGE and western blotting, with densitometric analysis
- structural and computational analysis: cryoSPARC, Phenix, FoldX, PDBePISA and AlphaFold
- molecular visualisation and graphics design (UCSF ChimeraX)
- use of Linux based high performance computing resources
- development and management of scientific social media channels for my research group

12. References

Agaronyan, K., Morozov, Y. I., Anikin, M., & Temiakov, D. (2015). Replication-transcription switch in human mitochondria. *Science, 347*(6221), 548–551.

Akbari, M., Visnes, T., Krokan, H. E., & Otterlei, M. (2008). Mitochondrial base excision repair of uracil and AP sites takes place by single-nucleotide insertion and long-patch DNA synthesis. *DNA Repair, 7*(4), 605–616.

Al-Behadili, A., Uhler, J. P., Berglund, A. K., Peter, B., Doimo, M., et al. (2018). A two-nuclease pathway involving RNase H1 is required for primer removal at human mitochondrial OriL. *Nucleic Acids Research, 46*(18), 9471–9483.

Alam, T. I., Kanki, T., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., & Kang, D. (2003). Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Research, 31*(6), 1640–1645.

Alexeyev, M., Shokolenko, I., Wilson, G., & LeDoux, S. (2013). The maintenance of mitochondrial DNA integrity—critical analysis and update. *Cold Spring Harbor Perspectives in Biology, 5*(5), a012641.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature, 290*(5806), 457–465.

Berk, A. J., & Clayton, D. A. (1974). Mechanism of mitochondrial DNA replication in mouse L-cells: asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. *Journal of Molecular Biology, 86*(4), 801–824.

Bogenhagen, D., & Clayton, D. A. (1978). Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence. *Journal of Molecular Biology, 119*(1), 49–68.

Bogenhagen, D. F., Rousseau, D., & Burke, S. (2008). The layered structure of human mitochondrial DNA nucleoids. *Journal of Biological Chemistry, 283*(6), 3665–3675.

Braithwaite, D. K., & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research, 21*(4), 787–802.

Brown, G. G., Gadaleta, G., Pepe, G., Saccone, C., & Sbisa, E. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *Journal of Molecular Biology, 192*(3), 503–511.

Calvo, S. E., Clauser, K. R., & Mootha, V. K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Research, 44*(D1), D1251–D1257.

Carrodeguas, J. A., Pinz, K. G., & Bogenhagen, D. F. (2002). DNA binding properties of human pol γB. *Journal of Biological Chemistry, 277*(51), 50008–50014.

Cazenave, C., Frank, P., Toulme, J. J., & Busen, W. (1994). Characterization and subcellular localization of ribonuclease H activities from Xenopus laevis oocytes. *Journal of Biological Chemistry, 269*(41), 25185–25192.

Cerritelli, S. M., & Crouch, R. J. (1998). Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase HI. *Genomics, 53*(3), 300–307.

Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS Journal, 276*(6), 1494–1505.

Cerritelli, S. M., Frolova, E. G., Feng, C., Grinberg, A., Love, P. E., & Crouch, R. J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Molecular Cell, 11*(3), 807–815.

Chan, S. S. L., & Copeland, W. C. (2009). DNA polymerase gamma and mitochondrial disease: understanding the consequence of POLG mutations. *Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1787*(4), 312–319.

Chang, D. D., & Clayton, D. A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proceedings of the National Academy of Sciences, 82*(2), 351–355.

Chang, D. D., Hauswirth, W. W., & Clayton, D. A. (1985). Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *The EMBO Journal, 4*(6), 1559–1567.

Clayton, D. A. (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annual Review of Cell Biology, 7*, 453–478.

Copeland, W. C. (2014). Defects of mitochondrial DNA replication. *Journal of Child Neurology, 29*(9), 1216–1224.

Cymerman, I. A., Chung, I., Beckmann, B. M., et al. (2008). EXOG, a novel paralog of Endonuclease G in higher eukaryotes. *Nucleic Acids Research, 36*(4), 1369–1379.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Engineering, Design and Selection, 3*(6), 461–467.

Doda, J. N., Wright, C. T., & Clayton, D. A. (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proceedings of the National Academy of Sciences, 78*(10), 6116–6120.

Dosekova, P., Dubiel, A., Karlowicz, A., Zietkiewicz, S., Rydzanicz, M., Habalova, V., Pienkowski, V. M., Skirkova, M., Han, V., Mosejova, A., Gdovinova, Z., Kaliszewska, M., Tońska, K., Szymanski, M. R., Skorvanek, M., & Ploski, R. (2020). Whole exome sequencing identifies a homozygous POLG2 missense variant in an adult patient presenting with optic atrophy, movement disorders, premature ovarian failure and mitochondrial DNA depletion. *European Journal of Medical Genetics*, *63*(4), 103821.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2022). Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021.10.04.463034.

Falkenberg, M., & Gustafsson, C. M. (2020). Mammalian mitochondrial DNA replication and mechanisms of deletion formation. *Critical Reviews in Biochemistry and Molecular Biology, 55*(6), 509–524.

Falkenberg, M., Larsson, N. G., & Gustafsson, C. M. (2024). Replication and Transcription of Human Mitochondrial DNA. *Annual Review of Biochemistry, 93*(1), 47–77.

Fan, L., Kim, S., Farr, C. L., Schaefer, K. T., Randolph, K. M., Tainer, J. A., & Kaguni, L. S. (2006). A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase. *Journal of Molecular Biology, 358*(5), 1229–1243.

Farge, G., & Falkenberg, M. (2019). Organization of DNA in mammalian mitochondria. *International Journal of Molecular Sciences, 20*(10), 2538.

Farge, G., Laurens, N., Broekmans, O. D., van den Wildenberg, S. M. J. L., Dekker, L. C. M., Gaspari, M., Gustafsson, C. M., Peterman, E. J. G., Falkenberg, M., & Wuite, G. J. L. (2012). Protein

sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nature Communications, 3*, 1013.

Farge, G., Mehmedovic, M., Baclayon, M., van den Wildenberg, S. M. J. L., Roos, W. H., Gustafsson, C. M., Wuite, G. J. L., & Falkenberg, M. (2014). In vitro-reconstituted nucleoids can block mitochondrial DNA replication and transcription. *Cell Reports, 8*(1), 66–74.

Farge, G., Pham, X. H., Holmlund, T., Khorostov, I., & Falkenberg, M. (2007). The accessory subunit B of DNA polymerase γ is required for mitochondrial replisome function. *Nucleic Acids Research, 35*(3), 902–911.

Fontana, G. A., & Gahlon, H. L. (2020). Mechanisms of replication and repair in mitochondrial DNA deletion formation. *Nucleic Acids Research, 48*(20), 11244–11258.

Fusté, J. M., Wanrooij, S., Jemt, E., Granycome, C. E., Cluett, T. J., et al. (2010). Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Molecular Cell, 37*(1), 67–78.

Gaidamakov, S. A., Gorshkova, I. I., Schuck, P., Steinbach, P. J., Yamada, H., Crouch, R. J., & Cerritelli, S. M. (2005). Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Research, 33*(7), 2166–2175.

Gillum, A. M., & Clayton, D. A. (1979). Mechanism of mitochondrial DNA replication in mouse L-cells: RNA priming during the initiation of heavy-strand synthesis. *Journal of Molecular Biology, 135*(2), 353–368.

González de Cózar, J. M., Carretero-Junquera, M., Ciesielski, G. L., Miettinen, S. M., Varjosalo, M., Kaguni, L. S., Dufour, E., & Jacobs, H. T. (2020). A second hybrid-binding domain modulates the activity of Drosophila ribonuclease H1. *Journal of Biochemistry, 168*(5), 515–533.

Gray, H., & Wong, T. W. (1992). Purification and identification of subunit structure of the human mitochondrial DNA polymerase. *Journal of Biological Chemistry, 267*(9), 5835–5841.

Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harbor Perspectives in Biology, 4*(9), a011403.

Greaves, L. C., Reeve, A. K., Taylor, R. W., & Turnbull, D. M. (2012). Mitochondrial DNA and disease. *The Journal of Pathology, 226*(2), 274–286.

Gustafson, M. A., Sullivan, E. D., & Copeland, W. C. (2020). Consequences of compromised mitochondrial genome integrity. *DNA Repair, 93*, 102916.

Gustafsson, C. M., Falkenberg, M., & Larsson, N.-G. (2016). Maintenance and expression of mammalian mitochondrial DNA. *Annual Review of Biochemistry, 85*, 133–160.

Hance, N., Ekstrand, M. I., & Trifunovic, A. (2005). Mitochondrial DNA polymerase gamma is essential for mammalian embryogenesis. *Human Molecular Genetics, 14*(13), 1775–1783.

Holmes, J. B., Akman, G., Wood, S. R., Sakhuja, K., Cerritelli, S. M., Moss, C., Bowmaker, M. R., Jacobs, H. T., Crouch, R. J., & Holt, I. J. (2015). Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication. *Proceedings of the National Academy of Sciences, 112*(30), 9334–9339.

Holt, I. J., Lorimer, H. E., & Jacobs, H. T. (2000). Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell, 100*(5), 515–524.

Holt, I. J. (2019). The Jekyll and Hyde character of RNase H1 and its multiple roles in mitochondrial DNA metabolism. DNA Repair, 84, 102630.

Hyjek, M., Figiel, M., & Nowotny, M. (2019). RNases H: Structure and mechanism. DNA Repair, 84, 102672.

Ikeda, M., Ide, T., Fujino, T., Arai, S., Saku, K., et al. (2015). Overexpression of TFAM or Twinkle increases mtDNA copy number and facilitates cardioprotection associated with limited mitochondrial oxidative stress. *PLOS ONE, 10*(3), e0119687.

Jemt, E., Persson, O., Shi, Y., Mehmedovic, M., Uhler, J. P., et al. (2015). Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Research, 43*(19), 9262–9275.

Johnson, A. A., & Johnson, K. A. (2001). Exonuclease proofreading by human mitochondrial DNA polymerase. *Journal of Biological Chemistry, 276*(41), 38097–38107.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.

Kaguni, L. S. (2004). DNA polymerase γ, the mitochondrial replicase. *Annual Review of Biochemistry, 73*, 293–320.

Kang, D., Miyako, K., Kai, Y., Irie, T., & Takeshige, K. (1997). In vivo determination of replication origins of human mitochondrial DNA by ligation-mediated polymerase chain reaction. *Journal of Biological Chemistry, 272*(24), 15275–15279.

Karlowicz, A., Dubiel, A. B., Czerwinska, J., Bledea, A., Purzycki, P., Grzelewska, M., McAuley, R. J., Szczesny, R. J., Brzuska, G., Krol, E., Szczesny, B., & Szymanski, M. R. (2022). In vitro reconstitution reveals a key role of human mitochondrial EXOG in RNA primer processing. *Nucleic Acids Research, 50*(14), 7991–8007.

Kaufman, B. A., Durisic, N., Mativetsky, J. M., Costantino, S., Hancock, M. A., Grutter, P., & Shoubridge, E. A. (2007). The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Molecular Biology of the Cell, 18*(9), 3225–3236.

Korhonen, J. A., Gaspari, M., & Falkenberg, M. (2003). TWINKLE has 5′->3′ DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *Journal of Biological Chemistry, 278*(48), 48627–48632.

Korhonen, J. A., Pande, V., Holmlund, T., Farge, G., Pham, X. H., Nilsson, L., et al. (2008). Structure-function defects of the TWINKLE linker region in progressive external ophthalmoplegia. *Journal of Molecular Biology, 377*(3), 691–705.

Korhonen, J. A., Pham, X. H., Pellegrini, M., & Falkenberg, M. (2004). Reconstitution of a minimal mtDNA replisome in vitro. *The EMBO Journal, 23*(12), 2423–2429.

Kornberg, A., & Baker, T. A. (1992). *DNA Replication*. W. H. Freeman.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology, 372(3), 774–797.

Kuhl, I., Miranda, M., Posse, V., Milenkovic, D., Mourier, A., et al. (2016). POLRMT regulates the switch between replication primer formation and gene expression of mammalian mtDNA. *Science Advances, 2*(9), e1600963.

Kukat, C., Davies, K. M., Wurm, C. A., Spåhr, H., Bonekamp, N. A., Kühl, I., Joos, F., Polosa, P. L., Park, C. B., Posse, V., Falkenberg, M., Jakobs, S., Kühlbrandt, W., & Larsson, N.-G. (2015). Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proceedings of the National Academy of Sciences, 112*(36), 11288–11293.

Kukat, C., Wurm, C. A., Spahr, H., Falkenberg, M., Larsson, N.-G., & Jakobs, S. (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and

frequently contain a single copy of mtDNA. *Proceedings of the National Academy of Sciences, 108*(33), 13534–13539.

Lee, S. R., & Han, J. (2017). Mitochondrial nucleoid: shield and switch of the mitochondrial genome. *Oxidative Medicine and Cellular Longevity, 2017*, 8060949.

Lee, Y. S., Kennedy, W. D., & Yin, Y. W. (2009). Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell, 139*(4), 828.

Lee, Y. S., Lee, S., Demeler, B., Molineux, I. J., Johnson, K. A., & Yin, Y. W. (2010). Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity. *Journal of Biological Chemistry, 285*(2), 1490–1499.

Li, Y., Liu, C., Jia, X., et al. (2024). RPA transforms RNase H1 to a bidirectional exoribonuclease for processive RNA–DNA hybrid cleavage. *Nature Communications, 15*, 7464.

Liang, X. H., Sun, H., Nichols, J. G., & Crooke, S. T. (2017). RNase H1-dependent antisense oligonucleotides are robustly active in directing RNA cleavage in both the cytoplasm and the nucleus. *Molecular Therapy, 25*(9), 2075–2092.

Longley, M. J., Nguyen, D., Kunkel, T. A., & Copeland, W. C. (2001). The fidelity of human DNA polymerase gamma with and without exonucleolytic proofreading and the p55 accessory subunit. *Journal of Biological Chemistry, 276*(41), 38555–38562.

Macao, B., Uhler, J. P., Siibak, T., Zhu, X., Shi, Y., et al. (2015). The exonuclease activity of DNA polymerase γ is required for ligation during mitochondrial DNA replication. *Nature Communications, 6*, 7303.

Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1678), 20140330.

McShane, E., Couvillion, M., Ietswaart, R., Prakash, G., Smalec, B. M., et al. (2024). A kinetic dichotomy between mitochondrial and nuclear gene expression processes. *Molecular Cell, 84*(8), 1541–1555.e9.

Mignotte, B., Barat, M., & Mounolou, J. C. (1985). Characterization of a mitochondrial protein binding to single-stranded DNA. *Nucleic Acids Research, 13*(5), 1703–1716.

Miralles Fusté, J., Shi, Y., Wanrooij, S., Zhu, X., Jemt, E., Persson, Ö., Sabouri, N., Gustafsson, C. M., & Falkenberg, M. (2014). In vivo occupancy of mitochondrial single-stranded DNA binding protein supports the Strand displacement mode of DNA replication. *PLoS Genetics, 10*(12), e1004832.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. Nature Methods, 19(6), 679–682.

Misic, J., Milenkovic, D., Al-Behadili, A., Xie, X., Jiang, M., Jiang, S., Filograna, R., Koolmeister, C., Siira, S. J., Jenninger, L., Filipovska, A., Clausen, A. R., Caporali, L., Valentino, M. L., La Morgia, C., Carelli, V., Nicholls, T. J., Wredenberg, A., Falkenberg, M., & Larsson, N. G. (2022). Mammalian RNase H1 directs RNA primer formation for mtDNA replication initiation and is also necessary for mtDNA replication completion. *Nucleic Acids Research, 50*(15), 8749–8766.

Moraes, C. T., Sciacco, M., Ricci, E., Tengan, C. H., Hao, H., Bonilla, E., Schon, E. A., & DiMauro, S. (1995). Phenotype-genotype correlations in skeletal muscle of patients with mtDNA deletions. *Muscle & Nerve, Suppl 3*, S150–S153.

Nicholls, T. J., & Minczuk, M. (2014). In D-loop: 40 years of mitochondrial 7S DNA. *Experimental Gerontology, 56*, 175–181.

Nowotny, M. (2009). Retroviral integrase superfamily: the structural perspective. *EMBO Reports, 10*(2), 144–151.

Nowotny, M., Cerritelli, S. M., Ghirlando, R., Gaidamakov, S. A., Crouch, R. J., & Yang, W. (2008). Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *The EMBO Journal, 27*(8), 1172–1181.

Nowotny, M., Gaidamakov, S. A., Ghirlando, R., Cerritelli, S. M., Crouch, R. J., & Yang, W. (2007). Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Molecular Cell, 28*(2), 264–276.

Oliveira, M. T., & Kaguni, L. S. (2011). Reduced stimulation of recombinant DNA polymerase γ and mitochondrial DNA (mtDNA) helicase by variants of mitochondrial single-stranded DNA-binding protein (mtSSB) correlates with defects in mtDNA replication in animal cells. *Journal of Biological Chemistry, 286*(47), 40649–40658.

Park, J., Baruch-Torres, N., & Yin, Y. W. (2023). Structural and molecular basis for mitochondrial DNA replication and transcription in health and antiviral drug toxicity. Molecules, 28(4), 1796.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1), 70–82.

Pham, X. H., Farge, G., Shi, Y., Gaspari, M., Gustafsson, C. M., & Falkenberg, M. (2006). Conserved sequence box II directs transcription termination and primer formation in mitochondria. *Journal of Biological Chemistry, 281*(34), 24647–24652.

Pohjoismäki, J. L., Holmes, J. B., Wood, S. R., Yang, M. Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., et al. (2010). Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *Journal of Molecular Biology, 397*(5), 1144–1155.

Posse, V., Al-Behadili, A., Uhler, J. P., Clausen, A. R., Reyes, A., Zeviani, M., Falkenberg, M., & Gustafsson, C. M. (2019). RNase H1 directs origin-specific initiation of DNA replication in human mitochondria. *PLoS Genetics, 15*(1), e1007781. https://doi.org/10.1371/journal.pgen.1007781

Posse, V., Shahzad, S., Falkenberg, M., Hallberg, B. M., & Gustafsson, C. M. (2015). TEFM is a potent stimulator of mitochondrial transcription elongation in vitro. *Nucleic Acids Research, 43*(5), 2615–2624.

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296.

Punjani, A., Zhang, H., & Fleet, D. J. (2020). Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods*, *17*(12), 1214–1221.

Punjani, A., & Fleet, D. J. (2021). 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology*, *213*(2), 107702.

Punjani, A., & Fleet, D. J. (2023). 3DFlex: Determining structure and motion of flexible proteins from cryo-EM. *Nature Methods*, *20*(6), 887–895.

Qiu, Y., Qian, P., Frank, U., Wintersberger, B., & Shen, B. (1999). Saccharomyces cerevisiae RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Molecular and Cellular Biology, 19*(12), 8361–8371.

Rahman, S., & Copeland, W. C. (2019). POLG-related disorders and their neurological manifestations. *Nature Reviews Neurology, 15*(1), 40–52.

Reyes, A., Kazak, L., Wood, S. R., Yasukawa, T., Jacobs, H. T., & Holt, I. J. (2013). Mitochondrial DNA replication proceeds via a ‘bootlace’ mechanism involving the incorporation of processed transcripts. *Nucleic Acids Research, 41*(11), 5837–5850.

Reyes, A., Melchionda, L., Nasca, A., Carrara, F., Lamantea, E., Zanolini, A., Lamperti, C., Fang, M., Zhang, J., Ronchi, D., Bonato, S., Fagiolari, M., Moggio, M., Ghezzi, D., & Zeviani, M. (2015). RNASEH1 mutations impair mtDNA replication and cause adult-onset mitochondrial encephalomyopathy. *The American Journal of Human Genetics, 97*(1), 186–193.

Riccio, A. A., Bouvette, J., Pedersen, L. C., Somai, S., Dutcher, R. C., Borgnia, M. J., & Copeland, W. C. (2024). Structures of the mitochondrial single-stranded DNA binding protein with DNA and DNA polymerase γ. *Nucleic Acids Research, 52(*17), 10329–10340.

Robberson, D. L., Kasamatsu, H., & Vinograd, J. (1972). Replication of mitochondrial DNA. Circular replicative intermediates in mouse L cells. *Proceedings of the National Academy of Sciences, 69*(3), 737–741.

Rosta, E., Yang, W., & Hummer, G. (2014). Calcium inhibition of ribonuclease H1 two-metal ion catalysis. Journal of the American Chemical Society, 136(8), 3137–3144.

Ruhanen, H., Ushakov, K., & Yasukawa, T. (2011). Involvement of DNA ligase III and ribonuclease H1 in mitochondrial DNA replication in cultured human cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1813*(11), 2000–2007.

Rubinstein, J. L., & Brubaker, M. A. (2015). Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of Structural Biology*, *192*(2), 188–195.

Rydberg, B., & Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proceedings of the National Academy of Sciences, 99*(26), 16654–16659.

Satoh, M., & Kuroiwa, T. (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research, 196*(1), 137–140.

Schon, E. A., Rizzuto, R., Moraes, C. T., Nakase, H., Zeviani, M., & DiMauro, S. (1989). A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. *Science, 244*(4902), 346–349.

Shutt, T. E., & Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics, 22*(2), 90–95.

Silva-Pinheiro, P., Pardo-Hernández, C., Reyes, A., Tilokani, L., Mishra, A., Cerutti, R., Li, S., Rozsivalova, D. H., Valenzuela, S., Dogan, S. A., Peter, B., Fernández-Silva, P., Trifunovic, A., Prudent, J., Minczuk, M., Bindoff, L., Macao, B., Zeviani, M., Falkenberg, M., & Viscomi, C. (2021). DNA polymerase gamma mutations that impair holoenzyme stability cause catalytic subunit depletion. *Nucleic Acids Research*, *49*(9), 5230–5248.

Sohl, C. D., Szymanski, M. R., Mislak, A. C., Shumate, C. K., Amiralaei, S., Schinazi, R. F., Anderson, K. S., & Yin, Y. W. (2015). Probing the structural and molecular basis of nucleotide selectivity by human mitochondrial DNA polymerase γ. *Proceedings of the National Academy of Sciences, 112*(28), 8596–8601.

Spelbrink, J. N., Li, F.-Y., Tiranti, V., Nikali, K., Yuan, Q.-P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.-M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H. T., Zeviani, M., & Larsson, C. (2001). Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nature Genetics, 28*(3), 223–231.

Spinelli, J. B., & Haigis, M. C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. *Nature Cell Biology, 20*(7), 745–754.

Sultana, A., & Lee, J. E. (2015). Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Current Protocols in Protein Science, 79*, 19.25.1–19.25.26.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., & Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science, 340*(6132), 619–621.

Suzuki, Y., Holmes, J. B., Cerritelli, S. M., Sakhuja, K., Minczuk, M., Holt, I. J., & Crouch, R. J. (2010). An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Molecular and Cellular Biology, 30*(21), 5123–5134.

Szczesny, B., Tann, A. W., Longley, M. J., Copeland, W. C., & Mitra, S. (2008). Long patch base excision repair in mammalian mitochondrial genomes. *Journal of Biological Chemistry, 283*(39), 26349–26356.

Szczesny, R. J., Hejnowicz, M. S., Steczkiewicz, K., Muszewska, A., Borowski, L. S., Ginalski, K., & Dziembowski, A. (2013). Identification of a novel human mitochondrial endo-/exonuclease Ddk1/c20orf72 necessary for maintenance of proper 7S DNA levels. *Nucleic Acids Research, 41*(6), 3144–3161.

Szymanski, M. R., Kuznetsov, V. B., Shumate, C., Meng, Q., Lee, Y. S., Patel, G., Patel, S., & Yin, Y. W. (2015). Structural basis for processivity and antiviral drug toxicity in human mitochondrial DNA replicase. *The EMBO Journal, 34*(14), 1959–1970.

Szymanski, M. R., Yu, W., Gmyrek, A., et al. (2017). A domain in human EXOG converts apoptotic endonuclease to DNA-repair exonuclease. *Nature Communications, 8*, 14959.

Tan, B. G., Mutti, C. D., Shi, Y., Xie, X., Zhu, X., Silva-Pinheiro, P., Menger, K. E., Díaz-Maldonado, H., Wei, W., Nicholls, T. J., Chinnery, P. F., Minczuk, M., & Falkenberg, M. (2022). The human mitochondrial genome contains a second light strand promoter. *Molecular Cell, 82*(19), 3646–3660.e9.

Tann, A. W., Boldogh, I., Meiss, G., Qian, W., Van Houten, B., Mitra, S., & Szczesny, B. (2011). Apoptosis induced by persistent single-strand breaks in mitochondrial genome: critical role of EXOG (5'-EXO/endonuclease) in their repair. *Journal of Biological Chemistry, 286*(37), 31975–31983.

Tiranti, V., Rocchi, M., DiDonato, S., & Zeviani, M. (1993). Cloning of human and rat cDNAs encoding the mitochondrial single-stranded DNA-binding protein (SSB). *Gene, 126*(2), 219–225.

Tsurumi, T., & Lehman, I. R. (1990). Release of RNA polymerase from vero cell mitochondria after herpes simplex virus type 1 infection. *Journal of Virology, 64*(1), 450–452.

Tyynismaa, H., Sembongi, H., Bokori-Brown, M., Granycome, C., Ashley, N., et al. (2004). Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number. *Human Molecular Genetics, 13*(24), 3219–3227.

Uhler, J. P., & Falkenberg, M. (2015). Primer removal during mammalian mitochondrial DNA replication. *DNA Repair, 34*, 28–38.

Uhler, J. P., Thörn, C., Nicholls, T. J., Matic, S., Milenkovic, D., Gustafsson, C. M., & Falkenberg, M. (2016). MGME1 processes flaps into ligatable nicks in concert with DNA polymerase γ during mtDNA replication. *Nucleic Acids Research, 44*(12), 5861–5871.

Wanrooij, P. H., Uhler, J. P., Simonsson, T., Falkenberg, M., & Gustafsson, C. M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proceedings of the National Academy of Sciences, 107*(37), 16072–16077.

Wanrooij, S., Fusté, J. M., Farge, G., Shi, Y., Gustafsson, C. M., & Falkenberg, M. (2008). Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proceedings of the National Academy of Sciences, 105*(32), 11122–11127.

Wanrooij, S., Miralles Fusté, J., Stewart, J. B., Wanrooij, P. H., Samuelsson, T., Larsson, N. G., et al. (2012). In vivo mutagenesis reveals that OriL is essential for mitochondrial DNA replication. *EMBO Reports, 13*(12), 1130–1137.

Wanrooij, P. H., Westerlund, F., Falkenberg, M., & Gustafsson, C. M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Research, 40*(20), 10334–10344.

Wong, L.-J. C., Naviaux, R. K., Brunetti-Pierri, N., Zhang, Q., Schmitt, E. S., Truong, C., Milone, M., Cohen, B. H., Wical, B., Ganesh, J., Basinger, A. A., Burton, B. K., Swoboda, K., Gilbert, D. L., Vanderver, A., Saneto, R. P., Maranda, B., Arnold, G., Abdenur, J. E., Waters, P. J., & Copeland, W. C. (2008). Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Human Mutation, 29*(9), E150–E172.

Wong, T. W., & Clayton, D. A. (1985). In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell, 42*(3), 951–958.

Wu, C.-C., Lin, J. L. J., Yang-Yen, H.-F., & Yuan, H. S. (2019). A unique exonuclease ExoG cleaves between RNA and DNA in mitochondrial DNA replication. *Nucleic Acids Research, 47*(10), 5405–5419.

Wu, C.-C., Lin, J. L. J., & Yuan, H. S. (2020). Structures, Mechanisms, and Functions of His-Me Finger Nucleases. *Trends in Biochemical Sciences, 45*(11), 935–946.

Wu, H., Lima, W. F., & Crooke, S. T. (1999). Properties of cloned and expressed human RNase H1. *Journal of Biological Chemistry, 274*(40), 28270–28278.

Xu, B., & Clayton, D. A. (1995). A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. *Molecular and Cellular Biology, 15*(2), 580–589.

Xu, B., & Clayton, D. A. (1996). RNA–DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA–DNA hybrids serving as primers. *The EMBO Journal, 15*(12), 3135–3143.

Yakubovskaya, E., Chen, Z., Carrodeguas, J. A., Kisker, C., & Bogenhagen, D. F. (2006). Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *Journal of Biological Chemistry, 281*(1), 374–382.

Yang, M. Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H. T., & Holt, I. J. (2002). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell, 111*(4), 495–505.

Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics, 44*(1), 1–93.

Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Molecular Cell, 22*(1), 5–13.

Yasukawa, T., Reyes, A., Cluett, T. J., Yang, M. Y., Bowmaker, M., Jacobs, H. T., & Holt, I. J. (2006). Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *The EMBO Journal, 25*(22), 5358–5371.

Yasukawa, T., & Kang, D. (2018). An overview of mammalian mitochondrial DNA replication mechanisms. *The Journal of Biochemistry*, 164(3), 183–193.

Ylikallio, E., & Suomalainen, A. (2012). Mechanisms of mitochondrial diseases. *Annals of Medicine, 44*(1), 41–59.

Young, M. J., Humble, M. M., DeBalsi, K. L., Sun, K. Y., & Copeland, W. C. (2015). POLG2 disease variants: analyses reveal a dominant negative heterodimer, altered mitochondrial localization and impaired respiratory capacity. *Human Molecular Genetics, 24*(18), 5184–5197.

Zhao, L., & Sumberaz, P. (2020). Mitochondrial DNA damage: prevalence, biological consequence, and emerging pathways. *Chemical Research in Toxicology, 33*(10), 2491–2501.