Atomistic Molecular Dynamics Simulations of Mitochondrial DNA Polymerase Gamma: Novel Mechanisms of Function and Pathogenesis

Liliya Euro [†,*,||], *Outi Haapanen* [‡,||], *Tomasz Róg* [‡,#], *Ilpo Vattulainen* [‡,#,%], *Anu Suomalainen* [†,&,!], *Vivek Sharma* [‡,#,$,*]

[†] Research Programs Unit, Molecular Neurology, University of Helsinki, 00290 Helsinki, Finland

[‡] Department of Physics, Tampere University of Technology, Tampere, Finland

[#] Department of Physics, University of Helsinki, Helsinki, Finland

[%] MEMPHYS-Center for Biomembrane Physics, University of Southern Denmark, Odense, Denmark

[&] Department of Neurology, Helsinki University Hospital, Helsinki, Finland

[!] Neuroscience Center, University of Helsinki, 00790 Helsinki, Finland

[$] Institute of Biotechnology, University of Helsinki, Helsinki, Finland

[||] Shared first author

Funding source statement:

VS and LE acknowledge funding from the Academy of Finland. AS has been supported by the Academy of Finland Centre of Excellence program, European Research Council, Sigrid Jusélius Foundation, Jane and Aatos Erkko Foundation, and University of Helsinki. IV thanks the Academy of Finland Center of Excellence program and the European Research Council (Advanced Grant project CROWDED-PRO-LIPIDS) for funding.

KEYWORDS: DNA replication, mitochondrial dysfunction, mitochondrial DNA replicase, molecular dynamics simulations

* Corresponding Authors: liliya.euro@helsinki.fi (LE), vivek.sharma@helsinki.fi (VS)

ABBREVIATIONS

Pol γ, heterotrimeric mitochondrial DNA polymerase gamma; Pol γA, the catalytic subunit of Pol γ; Pol γB, the accessory subunit of Pol γ; MIRAS, mitochondrial recessive ataxia syndrome; MD, molecular dynamics; PL, partitioning loop in Pol γA; IP, intrinsic processivity subdomain of the spacer domain in Pol γA; AID, accessory-interacting determinant subdomain of the spacer; NTD, N-terminal domain in Pol γA; exo, the exonuclease domain of Pol γA; pol, polymerase domain of Pol γ; ptDNA, primer-template DNA.

Abstract

DNA polymerase gamma (Pol γ) is a key component of the mitochondrial DNA replisome and an important cause of neurological diseases. Despite the availability of its crystal structures, the molecular mechanism of DNA replication, the switch between polymerase and exonuclease activities, the site of replisomal interactions and functional effects of patient mutations that do not affect direct catalysis have remained elusive. Here we report first atomistic classical molecular dynamics (MD) simulations of human Pol γ replicative complex. Our simulation data show that DNA binding triggers remarkable changes in the enzyme structure, including 1) completion of the DNA-binding channel via a dynamic subdomain, which in the apo-form blocks the catalytic site; 2) stabilization of the structure through the distal accessory β-subunit; and 3) formation of a putative transient replisome-binding platform in the "Intrinsic Processivity (IP)" subdomain of the enzyme. Our data indicate that non-catalytic mutations may disrupt replisomal interactions thereby causing Pol γ – associated neurodegenerative disorders.

DNA polymerases form a large class of enzymes catalyzing synthesis of new DNA strand complementary to the template DNA strand. Fourteen DNA polymerases have been described in humans, of which 13 function in nucleus [1]. The DNA polymerase gamma (Pol γ) is responsible for replication and repair of mitochondrial genome [1, 2], and its mutations are the most common Mendelian cause of mitochondrial disease in adults and children [3]. The manifestations of Pol γ dysfunction range from devastating infantile-onset Alpers-Huttenlocher syndrome, affecting the brain and the liver, to adult-onset progressive external ophthalmoplegia accompanied by either mtDNA depletions or multiple deletions (or both) in the affected tissues [4-9]. Some mutations directly affect the catalytic activities of Pol γ, but the two most common pathogenic mutations (A467T and W748S; amino acid numbering of A subunit), causing mitochondrial recessive ataxia syndrome (MIRAS, MSCAE) [4, 5, 10], have in *in vitro* studies shown no pronounced catalytic defect, thereby leaving their pathogenic molecular mechanism unexplained [10-12]. Altogether, Pol γ is an important protein in medicine, with an exceptionally high number of disease-associated variants – up to 250 (Human DNA Polymerase gamma Mutation Database, http://tools.niehs.nih.gov/polg/) [13].

Pol γ is a heterotrimer with one catalytic alpha subunit (Pol γA) and a homodimer of the accessory beta-subunit (Pol γB) [14]. Pol γA belongs to family A DNA polymerases and consists of N-terminal exonuclease (exo) and C-terminal polymerase (pol) domains connected by the spacer [2] (annotated protein sequence in Figure 1). The polymerase domain is further divided into thumb, fingers, and palm subdomains (Table S1). We previously showed that clustering of severe infantile disease mutations within the protein tertiary structure highlighted important functional domains, assigned as "Alpers clusters 1-5" [15]. Typically, the most severe diseases resulted from recessive mutations affecting two different Alpers clusters. Structural analysis of these five hot-spots of

natural mutagenesis, with synergistic effects concerning pathogenesis, provided new insight into the Pol γ molecular mechanism [15]. Our predictions concerning the side chain orientations of the amino acids surrounding the polymerase site, conformational changes in the fingers subdomain, and interaction of the Pol γA exo domain with the distal Pol γB upon DNA binding have all been corroborated by a recent crystallographic study of human Pol γ [16].

Pol γs have the highest fidelity among the known DNA polymerases [17, 18]. This precision is conferred by the high selectivity of the pol site and the high exonuclease activity of the enzyme. However, despite identification of several structures that are important for understanding the fidelity mechanism, the switch between the polymerase and exonuclease activities is still unclear [15, 16]. A structural analysis of the Alpers clusters suggested that a novel module named the "partitioning loop" (1049-1095 aa) is important for Pol γ fidelity [15]. Furthermore, modelling of human disease mutations in yeast suggested that a different fragment named the "orienter" is also part of the pol/exo switch mechanism [19]. A recent crystallographic study of human Pol γ proposed that the conserved β hairpin from the pol domain, which we previously called the RR-loop (842-856 aa) [15], is involved in communication between the pol and exo active sites [16]. Crystallographic studies and site-directed mutagenesis of *E. coli* DNA polymerase I (Klenow fragment), belonging to the same family A DNA polymerases, showed that the thumb subdomain and helix J (residues 858-861 in human Pol γA) are decisive for the orientation of the primer DNA towards either of the two catalytic sites (pol/exo) [20]. Additionally, a mode of interaction between the catalytic Pol γA and the distal accessory Pol γB subunits was proposed to be important for the regulation of Pol γ function [15, 21]. Taken together, these data bring out that Pol γ is a highly flexible enzyme, and despite a significant amount of work, its molecular mechanism is still unclear. Clearly, elucidating the mechanism would benefit if new approaches were available.

Here we approach this complex issue from a different perspective. To gain access to all the details in the machinery of Pol γ in a systematic manner, we perform atomistic molecular dynamics (MD) simulations of the replicative complex in conditions attempting to mimic its physiological environment. We show that both Pol γA and the distal Pol γB undergo structural re-arrangements upon loading onto the correctly paired primer-template DNA, involving all the aforementioned motifs. We find that all structural changes in Pol γ are driven and orchestrated by the structure of the binding DNA duplex on the side of the 3'-end of the growing strand. The results from MD simulations point to an unexpected function of the well-conserved partitioning loop. In addition, we observe that the binding of DNA in the Pol site induces structural re-organization in the intrinsic processivity (IP) subdomain of the spacer. The simulations of the human disease-causing Pol γ variants (A467T and W748S) indicate that these changes may disrupt a transient, putative replisome interaction domain, thereby providing a molecular understanding of their previously unknown pathogenic mechanism. In addition to these key results, there is reason to highlight the considerable added value that atomistic MD simulations can provide for experiments in this field. The method is versatile, being able to simultaneously address general as well as specific mechanistic questions concerning the role of individual amino acids on protein structure and function.

Materials and Methods

Molecular simulations provide atomistic insights into enzymatic processes that are difficult to achieve through experimental methodology [22]. Consistent with this view, MD simulations have been proved very useful in studying DNA polymerases, and provided various mechanistic insights that are difficult to obtain from experiments [23], and see references therein. Here, we report first fully atomistic (classical) MD simulations of Pol γ (PDB id: 3IKM [24]) in multiple states such as with and without DNA, and including also various *in silico* mutants (see Table 1).

Modelling of DNA on the Pol γ structure

We first structurally aligned the catalytic subunit of Pol γ (PDB id: 3IKM [24]) with the homologous subunit of T7 DNA polymerase (PDB id: 1T8E [25]), whose structure has been resolved in a DNA-bound form. After the structural alignment, we copied the DNA coordinates (from T7 DNA polymerase), and combined those with the apo Pol γ. The resulting DNA-bound Pol γ model displayed 'knots' (between DNA and protein) in certain regions, which were carefully removed manually, in order to minimize any perturbation in the protein structure. These changes involved translation of certain protein atoms by 0.5 – 1 Å to prevent the unphysical inter-locking of the protein with the DNA in the modelled system.

MD simulations

The constructed protein-DNA system was placed in a water-box (TIP3) [26], containing 100 mM Na^+ (or K^+) and 100 mM Cl^- as counterions to neutralize the charges as well as to model the conditions under physiological salt. The system with added potassium ions more closely mimics cellular conditions since cytoplasmic K^+ ion concentration is 100 mM versus 17-40 mM of Na^+,

whereas in mitochondria the concentration of K^+ ions is 150-180 mM versus 5-10 mM of Na^+ [27, 28]. There are various long and unresolved fragments in the crystal structure (see Table S1), which were not modelled, but the terminal amino acid residues were capped with charge-neutral patches. In order to remove steric clashes caused by unfavourable DNA-protein contacts, we first minimized and performed 20 ns equilibration run, in which harmonic restraints of 5 kcal mol^{-1} $Å^{-2}$ were imposed on the DNA heavy atoms. The restraints on DNA were necessary in order to prevent loss of base-pairing. Other approaches, in which protein was kept restraint (DNA unrestraint), were unsuccessful due to the rapid loss of hydrogen bonding interactions between the paired bases. The energy minimized DNA-protein model of Pol γ provided a reasonable starting point for longer simulations (see Figure S1). During the 20 ns run, the protein showed a maximum root mean square displacement (RMSD) of ca. 5.5 Å, in which the DNA-bound catalytic subunit showed the highest RMSD of ca. 6 Å, whereas accessory subunits were displaced by around 4.5 Å. All simulations were performed with the NAMD [29] program using the CHARMM27 force field for the protein, DNA, and salt ions [26, 30-32], and the Particle Mesh Ewald (PME) method was used for computing electrostatic interactions [33]. The cut-off for non-bonded interactions was 12 Å, and a smooth switching function was employed at 10 Å. In total, the model systems (Figure S2) comprised ca. 215,000 atoms. MD simulations were carried out at 310 K and 1 atm pressure, and up to 1 μs per simulation (Table 1). The Langevin temperature [34], and Nosé Hoover Langevin pressure [35, 36] control methods, as implemented in NAMD, were applied. The time step used in integrating the equations of motion was 1 fs.

Construction of multiple independent model systems

For all model systems, multiple independent simulations considered in this study were performed in order to obtain sampling/statistics (see Table 1 for the various model systems and their replicas). In order to understand if the secondary structure of the partitioning loop is stable, and to consider how it evolves upon DNA binding, we performed four independent simulations. In each of the four replicas (model system # 3), a small loop (residues 1060 - 1064) was translated away from the DNA since it clashed with the DNA structure. The four systems were then subjected to minimization and MD simulations as follows; in replica # 1, 10000 steps minimization, a 20 ns MD run with 5 kcal $mol^{-1}Å^{-2}$ harmonic restraints on DNA, followed by a 400 ns production run; in replica # 2, 10000 steps minimization, 1 ns MD run with 5 kcal $mol^{-1}Å^{-2}$ harmonic restraints on paired DNA and on short helix 1053 – 1059 of the partitioning loop, followed by a production run; during the entire production run of replica # 3, the DNA and helices of the loop (1039 - 1050 and 1053 - 1059) were subjected to harmonic restraints of 5 kcal $mol^{-1}Å^{-2}$; and in replica # 4, prior to the production run, the unpaired template DNA tail was cleaved, system was minimized for 10000 steps, followed by a 10 ns of MD run with 5 kcal $mol^{-1}Å^{-2}$ harmonic restraints on DNA.

The A467T, E1143G/W748S and triple-Ala *in-silico* mutants (Table 1) were constructed from the above partitioning loop model system (see the fourth partitioning loop-replica) by modifying the relevant coordinate/structure files. This was followed by a 10000 step minimization, up to a 20 ns MD run with restrained DNA (5 kcal $mol^{-1}Å^{-2}$) and production runs (400 ns). For the triple-Ala mutant, the second replica was initiated from the last snapshot of the 1 μs DNA+ run (model system # 2 in Table 1). The simulations of mutants (model systems 4-6) and different conformations of the partitioning loop (system 3) were discontinued after it was apparent that no major structural transitions were taking place.

During the completion of this study, a new structure of Pol γ with a bound substrate DNA (PDB id: 4ZTU/4ZTZ [16]) was reported. Therefore, this structure was also considered in this work (Figure S3), and additional simulations were performed (see Table 1).

Simulation analysis

The VMD [37] program was used for analysing the simulation trajectories. For detailed analysis of specific structural changes caused by the binding of DNA, snapshots taken from simulation trajectories were also analyzed by the Accelrys Discovery Studio v4.1 package. Due to multiple unresolved fragments in the reported crystal structures, and certain large scale conformational changes, we observed high root mean square displacement (RMSD) in our simulations (Figure S4). Nevertheless, the simulation convergence [38] is apparent from the stabilization of the RMSD in all long-timescale simulations (see Figure S4). Moreover, the two-dimensional (2D) RMSD plots of the catalytic subunit (Figure S5) and the entire protein (Figure S6) reveal that the DNA-bound protein simulations have achieved convergence at around 400 ns, and in the latter half of the simulation (400 – 1000 ns) no major conformational changes took place. In contrast, the simulation of the protein devoid of DNA was probably not converged, and was likely to visit multiple conformational sub-states, including the ones more likely to bind DNA.

The three-subunit enzyme (PDB id: 3IKM) is split into 10 separate fragments due to missing electron density in the crystal structure (see Table S1). This likely caused larger fluctuations in the protein structure (see Figures S4 and S7). We also calculated the RMSD of three subdomains of Pol γ (exo, IP, and AID subdomains, see also Table S1) by aligning these segments individually. The plots revealed that the RMSD is indeed lower when compared to the RMSD

calculated based on alignment of the entire protein (see Figure S7). The RMSD plots of all other simulations are provided in Figure S8.

Results

To study how the binding of Pol γ with the correctly paired DNA duplex re-organizes the protein structure, we performed MD simulations of the replicase with and without primer-template-DNA (ptDNA) residing in the pol site (Table 1). This approach allowed us to spot fragments in the Pol γ structure, which undergo conformational re-arrangements in response to DNA binding (Table S2, and Figures 1 and 2), previously unpredicted by crystallography [16,24].

Completion of the DNA-binding channel by the partitioning loop (PL) and the Orienter

Comparative analyses of the 1 μs MD simulations of Pol γ with and without DNA reveal detailed dynamics of the formation of DNA-binding channel, implicating that the channel is formed to completion only after DNA binding. We observe that the insertion of ptDNA into the pol site causes partial unwinding of the 799-809 aa fragment in the thumb subdomain (Figure S9 panel A), and minor conformational changes also in the palm 858-868 aa fragment. Both of these fragments are in close proximity to the pol site and the RR-loop (Figures 2 and 3). The S799-F800-W801 motif with a downstream stretch of residues 802-807 (labeled "e" in Figure 1), and the fragment 858-861 (labeled "f" in Figure 1, corresponding to the J-helix in the Klenow fragment [20]) also reorganize rapidly upon optimal binding and restraining of ptDNA in the pol site (Figure 3). The partial destabilization of the helix in the thumb subdomain is also observed in the new crystal structure of DNA-bound Pol γ [16]. The same scenario is also supported when independent MD simulations are performed in KCl as salt solution (see panels A, B, E and F in Figure S9).

Our analysis of the simulation data show that the re-arrangements of the thumb subdomain and the Pol fragment "f" are synchronized by bound DNA. The Pol "f" fragment and the "SFW segment" (SYW in *D. melanogaster*) were previously studied in bacterial Pol I and drosophila,

respectively [20, 39]. The SYW triple-alanine mutant was proposed to have "antimutator" characteristics, causing a dramatic reduction in the DNA-binding and polymerase activity, but an increase in the exonuclease activity [39]. The same effect was shown for P680 (A865 in human Pol γA) from the J-helix in the Klenow fragment [20]: P680G/Q perturbed the structure and the location of the J-helix, and resulted in a switch into the exonuclease mode [20]. These data suggested that this specific protein region is crucial for the exonuclease switch. We therefore generated a triple-alanine mutation of the SFW motif *in silico* and performed two independent 400 ns simulations of the resulting mutant with inserted DNA (model system 5 in Table 1, see also the Methods section for differences in the two replicas). These data show that in the SFW mutant the DNA-protein interface becomes loose, while in the wild-type (WT) protein it remains relatively rigid (see Figures 3 and 4). The same distances between the protein and the DNA also remain roughly similar in the 400 ns WT Pol γ simulation in KCl as salt; distances between the nucleic acid base 1020 and the residues 799, 800, and 801 are 7.11 Å ± 0.31, 4.56 Å ± 0.27 and 7.47 Å ± 0.34, respectively. Based on these results, we propose that the increased structural flexibility in the entrance of the DNA-binding channel prevents formation of a stable polymerization complex with bound DNA and promotes shifting of the enzyme into the exo mode, thereby explaining the observed weakened DNA binding and the enhanced exonuclease activity of the "antimutator" J-helix and the SYW triple alanine mutant.

The simulations reveal functionally important dynamic conformational changes in the DNA-binding channel involving the orienter (304-312 aa) and the partitioning loop (1044-1095 aa) (Figures 1 and 2). The orienter is part of the exo domain, but in the tertiary structure it forms a loop-helix motif and is positioned close to the pol catalytic site (Figures 1, 2). The known disease mutations in this short helix (L304R, S305R, Q308H, R309L/H/C, and W312R) associate with

severe recessive ataxia/epilepsy disorders, and result in an increased mutation rate, decreased DNA binding, and a lowered pol activity [40]. In yeast Pol γ, L304R, R309L/H, and W312R mutants, similar to the SFW mutant, showed increased exo activity, suggesting interaction of this fragment with DNA in the replicative mode [19]. Our simulation data show alignment of the orienter helix along the upstream DNA duplex (Figure 5), and formation of multiple interactions between residues S305, S306, F307, Q308 and the first three upstream bases of template DNA (Figure 5B, inset), in agreement with the structure of the recently solved ternary complex [16].

Partitioning loop is an insertion in the fingers subdomain, existing exclusively in the Pol γ group of enzymes. The structure of the apoenzyme showed that it consists of two short helices (1053-1058 aa and 1092-1100 aa), separated by a loop (1059-1073 aa) and a crystallographically unresolved part (1075-1089 aa) [24]. Structural studies also suggested that some parts of this unique fragment are highly flexible and block the entrance of the DNA-binding channel from the pol site, requiring it to relocate upon loading of substrate DNA [24]. Data from multiple independent simulations of Pol γ with bound DNA (Table 1) show that the domain (1044-1100 aa) containing the partitioning loop loses its secondary structure, and gradually aligns along the solvent exposed side of upstream ptDNA, and partially parallel to the orienter segment (Figures 2 and 5B). This is also displayed as loss of helicity in the segment 1053-1058 of the partitioning loop (see Figures S10, and also S3). In contrast, MD simulations of the enzyme without DNA (model system 1) show that partitioining loop retains its crystallographic conformation (Figures 5A and S10). The same scenario is also observed when simulations are performed in KCl as salt solution (model systems 1^a and 2^a, see Figure S10). These data suggest that the binding of DNA rearranges the Pol γ structure to form a completed DNA-binding channel and to physically restrain DNA for efficient polymerization (Figures 2 and 5). Our data also agree well with the recent crystal structure data [16]

that multiple conformational sub-states are possible for the partitioning loop. In the crystal structure of DNA-bound form of Pol γ [16], the loop attains an entirely different conformation than in the resolved apo form (see also Figure S3). Furthermore, our results provide a mechanism to explain how mutations in this loop, by perturbing protein-DNA contacts, result in a similar phenotype as the orienter mutations, the SYW triple alanine mutant, and the affected pol "f" fragment (J-helix).

Further dynamic adjustments in the distal part of the DNA-binding surface, formed by the Intrinsic Processivity and AID (accessory-interacting determinant) subdomains of the spacer, are ascertained upon analysis of simulation trajectories. Sequence stretches "a" (499-504 aa or K-tract), "b" (583-586 aa), "c" (755-759 aa), and "d" (768-770 aa), assigned according to their order in the Pol γA sequence, gradually align along the upstream DNA duplex, and form crests directed towards its minor and major grooves (Figures 1 and 2). These multiple interactions translate to tight grip on DNA, enabling the maintenance of the optimal tilt angle and consequent maximum polymerization rate. Importantly, we point out that all the DNA-binding fragments, revealed by our simulations, overlap with those in the crystal structure of Pol γ with bound DNA [16], but provide additional understanding of dynamics, DNA-dependence and flexibility. The overlaps however, further validate the accuracy and biological relevance of our atomistic simulation results.

Exonuclease domain rotates upon formation of the replicative complex

The recent crystal structure of the Pol γ DNA-bound ternary complex displayed sparse interactions between the exo domain and the bound substrate DNA [16] (see Figure S11, panel A). Consistent with this view, our MD simulations on DNA-bound Pol γ show that the entire exo domain rotates away from the substrate DNA (video S1), and at the end of the simulation resides in a position

similar to that observed in the crystal structure of the ternary complex (Figure S11, panel B). This rotational movement stands in strong contrast to the data from the DNA-free simulation, in which the exo domain retains its crystallographic conformation (video S1). The rotational movement of the exo domain away from the bound-DNA is also shown as a distance plot between the Cα of functionally important R227 and the nearest DNA backbone atom (Figure 6). The distance increases and relaxes to ca. 45 Å, which is similar to what is observed from the control simulations of DNA-bound Pol γ crystal structure (Figure 6). We suggest that the rotated conformation, with a changed mutual disposition of pol and exo active sites, likely represents the polymerization mode of Pol γ.

Distal Pol γB stabilizes spacer and the exo domain of the catalytic subunit in the replicative mode

In the crystal structure of the DNA-bound ternary complex, the entire distal Pol γB rotated towards the exodomain of the catalytic subunit [16] and formed several non-covalent bonds with the residues on the tip of the exo domain close to R232, in agreement with an earlier proposal [15]. However, our simulation data show that interactions between Pol γA and the distal accessory Pol γB are likely to be more complex and dynamic. The first interface is formed between the AID subdomain fragment encompassing the residues 524-542 aa (assigned as "A3" in Figure 2), the helix 246-264 aa, and the loop 458-463 aa (fragments "B1" and "B4", respectively, in Figure 2) in distal Pol γB. This interaction stabilizes the distal part of the DNA-binding channel (Figures 1, 2), as also shown by formation of a larger number of contacts between the two domains in different simulations (Figure S11, panels C and D). The second inter-subunit contact surface is formed between the loops 203-207 aa and 376-383 aa ("A1" and "A2" in Figure 2, in respective order) in catalytic Pol

γA, and between the loops 388-395 aa and 444-451 aa in distal Pol γB ("B2" and "B3" in Figure 2, respectively), stabilizing the new position of the exo domain, likely adapted for the polymerization mode.

DNA binding triggers formation of the replisome-binding site within the Intrinsic Processivity subdomain

The Intrinsic Processivity subdomain harbors those human disease variants that have primarily nervous system and liver phenotypes, differing from the catalytic-site mutations that manifest often as myopathy and parkinsonism [3]. However, the pathogenic consequences of these mutations for Pol γ have remained elusive, as they do not affect the catalytic functions. Our 1 μs MD simulation of DNA-bound form of Pol γ shows that the surface of the Intrinsic Processivity subdomain in the spacer region undergoes a conformational change upon DNA binding, involving the fragments 618-673 aa and 723-750 aa (marked with purple and dark blue in Figures 1 and 2, respectively). In the course of the simulation, the loop G723-P745 rotates 180° away from the pol site and the bound DNA, and induces formation of a stable surface exposed β-sheet-cluster composed of two short β-strands G619-L623 and G746-F750 (Figures 7B and S12, and video S2). This observed 180° rotation of the loop is also observed when an additional simulation is performed in KCl as salt solution. The flipping of the loop away from bound-DNA is displayed as a distance plot between the Cα atom of the loop residue His733 and the P atom of the nucleic base 1017 in two different simulation conditions (Figure 8). The current data, and the missing electron density (of flipping loop residues 723-736) in this region in the recent crystal structure of DNA-bound enzyme strongly advocates for the highly dynamic Intrinsic Processivity subdomain [16] (see also Figure S3). Moreover, congregation of various known pathogenic mutations without catalytic phenotypes

(L623W, R627W/Q, P648R, G746S, W748S, and F749S) within this β-cluster also argues for its functional role. Together the evidence suggests that the β-cluster and the region around it form a tentative replisome protein interaction site.

We next analyzed the core region of the Intrinsic Processivity subdomain, and found it to become more compact upon loading on DNA, forming dense hydrophobic interactions connecting the β-cluster with the upper tip of the thumb (Figure S13). The Cα-Cα distance between W748 and A467, the sites of the two most common human disease mutations [4, 41] located on the opposite edges within the Intrinsic Processivity core, fluctuates around 22 Å in the 1 μs simulation of the ternary complex, whereas in the DNA-free form this distance largely exceeds the crystallographic (PDB id: 3IKM) value of ca. 25 Å (Figure S13, and panel A in Figure S14). Similarly, in the simulations performed with different salt species (KCl) the distance increases to ca. 35 Å, whereas in the DNA-bound form the distance starts to relax to much lower values (panel B in Figure S14).

The hydrophobic interaction network within the Intrinsic Processivity subdomain core focuses on the invariant L602, and involves L466, C471, W486, V598, Y614, S615, F766, A786, and A791. The solvent accessible surface area (SASA) calculation on this cluster reveals that it is tightly packed in WT DNA-bound polymerase in agreement with the distance analysis (Figure S14, panels A and B). The SASA for model systems 1 and 1^a are 2237 ± 84 $Å^2$ and 2282 ± 79 $Å^2$, and for systems 2 and 2^a are 1867 ± 96 Å and 2208 ± 78 $Å^2$, respectively. Establishment of the highly packed hydrophobic bulk of the Intrinsic Processivity subdomain upon DNA binding likely serves two functions: 1) promotes the flipping of the loop surrounding the β-cluster, and stabilizes the β-cluster at the putative replisome-binding site (see Figures 8 and S12), and 2) indirectly assists DNA binding by restructuring the fragment 799-809 aa of the thumb unit (fragment "e" in Figure 1, see also Figure S9), which is critical to restrain DNA in the replicative ternary complex (Figure

S15). Furthermore, our results allow us to explain the *in vitro* mutagenesis data of the key leucine 602, which after being probed in the fly enzyme (L558 in *D. melanogaster*) as a part of the triple alanine mutant of P556A/K557A/L558A resulted in complete loss of DNA binding and pol activity [39].

Recessive ataxia mutations, W748S and A467T, perturb formation of the replisome-binding site

Homozygous disease variants W748S (always co-occurring allelic with the E1143G variant) and A467T underlie phenotypically similar disorders, and at a molecular level, are accompanied by mtDNA instability in affected tissues [10, 12]. Structural analysis of W748S using DNA free Pol γ structure predicted no direct effect on the pol catalytic activity, whereas analysis of E1143G was predicted to have a modest effect on polymerase rate [15]. The A467T mutation was predicted to affect the hydrophobic core of the spacer at the interface between the Intrinsic Processivity, AID and the thumb subdomains [15] - the area, which is critical for the formation of the DNA-binding channel. Since both W748 and A467 are located at this structurally dynamic area, it is imperative to ask how W748S and A467T mutants would affect the structural re-arrangements in the Intrinsic Processivity subdomain. To understand this, we separately modeled and simulated W748S (together with E1143G) and A467T mutations *in silico* using Pol γ with bound DNA (see Table 1). This analysis revealed that W748S and A467T perturb the density of the stabilizing hydrophobic core in the Intrinsic Processivity subdomain, as shown by an increase in the distance between W748 and A467 in mutant simulations (compare panels C and D with panels A and B in Figure S14), and SASA calculations; for two replicas of model systems 4 and 6, the SASA values obtained are 2256 ± 48/2362 ± 44 Å^2 and 2238 ± 42/2123 ± 96 Å^2, respectively. This also makes

it insufficient to support the formation of a stable surface exposed β-cluster and promote the unwinding of the thumb subdomain for DNA binding (see Figures 7, S12 and S15). In contrast, E1143G substitution did not result in any considerable local structural changes that would affect the function of the replicative complex. Also, a hydrophilic and smaller residue such as serine in the position 748 is likely to destabilize the local interactions with S615 and Y614, which are part of the extended interaction network around L602 (Figure S13). These findings suggest that upon DNA binding both W748S and A467T perturb the replisome-protein interactions, giving rise to the observed non-catalytic phenotypes.

We found in two independent 400 ns simulations that the substitution of alanine 467 for a much more hydrophilic threonine reduces the hydrophobicity close to L602 (Figure S16 panel C), followed by remodeling of the intra-subunit hydrophobic bonding network around this residue (Figure S13). The resulting loosened core of the IP subdomain (Figure S14, panels C and D) in both mutants proposes its inability to support the formation of the β-cluster exposed on the Intrinsic Processivity surface upon DNA binding (Figures 7 and S12 panels E-H). Structural changes around L602 due to mutations also affect the thumb and its mode of DNA binding. For instance, in both replicas of A467T and in one replica of W748S mutants, the fragment "e" in the thumb subdomain preserves its helical structure despite the bound DNA (panels G-I in Figure S9 and Figure S15). The A467T mutation has no effect on the interaction with the distal accessory subunit, which stabilizes the distal half of the DNA-binding channel in Pol γA and compensates for the absence of a tight grip and fine alignment of ptDNA by the thumb close to the pol site. This may explain the alleviation of the mutation effect by Pol γB *in vitro* [10].

Discussion and Concluding Remarks

We report here atomistic MD simulations to reveal the structure, dynamics, and molecular mechanism of the human mitochondrial replicative polymerase, Pol γ, in its polymerization mode, with important mechanistic implications for human disease mutations. These data, combined with knowledge from previous structural and biochemical studies, enhances our understanding of the functional characteristics of this important protein, in which more than two hundred mutations have been reported to underlie variable forms of human disease.

We performed ca. 8.5 μs simulations in full atomic detail to consider Pol γ in a variety of different conditions with and without bound DNA, and with and without specific mutations. To determine mechanistic details, we used the structure of the DNA free form of Pol γ (PDB id: 3IKM [24]) and modeled conformational changes in protein structure that take place upon loading onto the correctly paired DNA substrate. Overall, the observed structural changes replicate most of those found in the recently resolved Pol γ ternary structure [16] and support our earlier predictions [15]. However, the atomistic MD simulations also reveal important dynamic aspects of the enzyme that are relevant for its functions and associated diseases.

Our simulation results suggest that the function of the partitioning loop in the replicative complex is to complete the DNA-binding channel. In the structural study of DNA-free Pol γ, the partitioning loop blocks the pol site, and a simple rotation was suggested for it to enable DNA binding in the catalytic site [24]. We previously proposed, based on the molecular phenotype caused by the mutations in this loop that it contributes to the high-fidelity of Pol γ [15]. Unexpectedly, in the recent structure of DNA-bound Pol γ [16] the partitioning loop (fragment 1052-1086 aa) displaced upstream sequence stretch 995-1030 and accommodated its place on the protein surface (Figure S3). Although these data indicated a high conformational flexibility of the partitioning loop, they provided no functional explanation of why many pathologic mutations are located in this segment.

Mutations R1047W and P1073L are found in early-childhood Alpers syndrome, whereas G1051R and G1076V have been reported to associate with a late-onset progressive external ophthalmoplegia (Human DNA Polymerase gamma Mutation Database, http://tools.niehs.nih.gov/polg/). Mutations G1051R, P1073L, and G1076V increased mutation frequency of DNA synthesis [40, 42], resembling the phenotype of mutations in the orienter domain [19]. In the new crystal structure, the amino acids G1051 and R1047 showed no interactions with the rest of the protein, whereas G1076 occupied the same place as I1020 in the apo form, thereby questioning the pathogenicity of valine in this position. Analysis of Pol γ crystal structures with and without bound DNA suggested that the partitioning loop is highly flexible [16, 24]. Indeed, in our MD simulations of the replicating complex, this loop partially unwound and aligned along DNA, thereby contributing to physical restraining of the substrate in the pol mode. Importantly, this unique conformation also allows us to explain the pathogenicity of mutations located in the partitioning loop.

One interesting observation from our simulation data is the structural rearrangement of the Intrinsic Processivity subdomain. The recent structure of the Pol γ DNA-bound ternary complex [16] showed that the Intrinsic Processivity subdomain is indeed highly dynamic, as it was more disordered in the ternary complex than in the DNA-free form (Figure S3). Consistent with this, MD simulations revealed that upon DNA-binding the β-sheet cluster within the Intrinsic Processivity subdomain is stabilized and becomes more solvent exposed. We suggest that *in vivo* this part of the spacer represents the replisome contact site and enables interaction with the helicase TWINKLE and/or the single-strand binding protein (mtSSB). Our hypothesis is also strengthened by the fact that all known non-catalytic pathogenic mutations co-localize within this cluster. In our previous work [15], we assigned the IP-subdomain mutations as "Alpers cluster 5" mutations

affecting the replisomal interactions. Our assessment was based on biochemical data from site-directed mutagenesis of the G575-W576-F578 motif in fly Pol γ [39], corresponding to G619, W620, and Y622 in human, and locating within one of the strands in the β-sheet cluster (Figure 7A). Substitution of each residue from the GWF motif for alanine resulted in a loss of activity stimulation by mtSSB [39]. Our simulation data show that upon DNA binding almost all "Alpers cluster 5" residues, together with the GWY motif, locate in the tentative binding site for replisome proteins. The transient nature and DNA-binding-dependent nature of the docking site may also explain why trials to co-precipitate Pol γ with Twinkle and SSB have been unsuccessful.

The presence of a transient replisome-binding site may also provide insights into the pathogenic effects of non-catalytic mutations W748S and A467T. The two mutants perturb the formation of a stable β-cluster in the IP subdomain. Therefore, the inability of the two mutants to bind TWINKLE would reduce the rate of replication initiation, and processive mtDNA unwinding catalyzed by the helicase. Additionally, both mutations would render Pol γA non-responsive for mtSSB stimulation, leading to a higher rate of polymerase dissociation from its substrate and mtDNA depletion, thereby leading to the observed disease phenotypes.

The data from MD simulations rationalize the necessity of the two accessory Pol γB subunits. Whilst the proximal one enhances DNA binding by the catalytic subunit and brings together catalytic and distal Pol γB, the distal one stabilizes the AID subdomain upon formation of the DNA-binding channel, and assists the rotation of the exo domain within the replicating complex. In this context, it is puzzling how the functions of distal Pol γB are accomplished in the fly Pol γ holoenzyme, containing only one accessory subunit [43]. MD simulations also provide a mechanistic explanation for the results of previous site-directed mutagenesis studies of T440A, T447A, and K463A in human Pol γB, which resulted in decreased DNA binding of the holoenzyme

[44]. These mutations may perturb the interactions of the fragments "B3" and "B4" in the distal Pol γB with the catalytic subunit, thereby destabilizing the distal part of DNA-binding channel. The atomistic simulations also shed light on the molecular pol/exo switch mechanism of Pol γ. Our data indicates that the high rate and the fidelity of DNA replication rely on the tight encircling of the first four base pairs of ptDNA within the binding channel of the enzyme. This binding is a highly coordinated and concerted action and completed by the structural changes induced by DNA binding within several separate structural fragments contributing to the channel, and spatially placed in the immediate vicinity of the pol site.

In conclusion, we demonstrate here the potential of multidisciplinary evaluation of protein structures, combining knowledge from disease phenotypes, biochemical data and static crystal structure analysis, combined with dynamic atomistic MD simulations, in providing a high-gain framework for functional assessment of enzymes, or their selected sequence fragments or residues. Furthermore, we show the power of molecular dynamics simulation approach, allowing analysis of protein states and structures beyond resolution capacity of current experimental methodology.

ACKNOWLEDGMENT

We acknowledge CSC – IT Center for Science (Espoo, Finland) for excellent computing resources, and access to using the Accelrys Discovery Studio package.

SUPPORTING INFORMATION AVAILABLE

Figures of the constructed model system, simulation trajectory analysis and simulation snapshots are provided in the supporting information file. Also, description of various protein subdomains and structural elements studied in this work are summarized in the tables in the supporting information file. This material is available free of charge via the Internet at http://pubs.acs.org.

References

[1] Graziewicz, M. A., Longley, M. J., and Copeland, W. C. (2006) DNA polymerase gamma in mitochondrial DNA replication and repair, *Chem Rev 106*, 383-405.

[2] Kaguni, L. S. (2004) DNA polymerase gamma, the mitochondrial replicase, *Annu Rev Biochem 73*, 293-320.

[3] Ylikallio, E., and Suomalainen, A. (2012) Mechanisms of mitochondrial diseases, *Ann Med 44*, 41-59.

[4] Hakonen, A. H., Heiskanen, S., Juvonen, V., Lappalainen, I., Luoma, P. T., Rantamaki, M., Goethem, G. V., Lofgren, A., Hackman, P., Paetau, A., Kaakkola, S., Majamaa, K., Varilo, T., Udd, B., Kaariainen, H., Bindoff, L. A., and Suomalainen, A. (2005) Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin, *Am J Hum Genet 77*, 430-441.

[5] Suomalainen, A., and Isohanni, P. (2010) Mitochondrial DNA depletion syndromes--many genes, common mechanisms, *Neuromuscul Disord 20*, 429-437.

[6] Suomalainen, A., Majander, A., Wallin, M., Setala, K., Kontula, K., Leinonen, H., Salmi, T., Paetau, A., Haltia, M., Valanne, L., Lonnqvist, J., Peltonen, L., and Somer, H. (1997) Autosomal dominant progressive external ophthalmoplegia with multiple deletions of mtDNA: clinical, biochemical, and molecular genetic features of the 10q-linked disease, *Neurology 48*, 1244-1253.

[7] Van Goethem, G., Luoma, P., Rantamaki, M., Al Memar, A., Kaakkola, S., Hackman, P., Krahe, R., Lofgren, A., Martin, J. J., De Jonghe, P., Suomalainen, A., Udd, B., and Van Broeckhoven, C. (2004) POLG mutations in neurodegenerative disorders with ataxia but no muscle involvement, *Neurology 63*, 1251-1257.

[8] Winterthun, S., Ferrari, G., He, L., Taylor, R. W., Zeviani, M., Turnbull, D. M., Engelsen, B. A., Moen, G., and Bindoff, L. A. (2005) Autosomal recessive mitochondrial ataxic syndrome due to mitochondrial polymerase gamma mutations, *Neurology 64*, 1204-1208.

[9] Zeviani, M., Servidei, S., Gellera, C., Bertini, E., DiMauro, S., and DiDonato, S. (1989) An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region, *Nature 339*, 309-311.

[10] Luoma, P. T., Luo, N., Loscher, W. N., Farr, C. L., Horvath, R., Wanschitz, J., Kiechl, S., Kaguni, L. S., and Suomalainen, A. (2005) Functional defects due to spacer-region mutations of human mitochondrial DNA polymerase in a family with an ataxia-myopathy syndrome, *Human molecular genetics 14*, 1907-1920.

[11] Palin, E. J., Lesonen, A., Farr, C. L., Euro, L., Suomalainen, A., and Kaguni, L. S. (2010) Functional analysis of H. sapiens DNA polymerase gamma spacer mutation W748S with and without common variant E1143G, *Biochim Biophys Acta 1802*, 545-551.

[12] Hakonen, A. H., Goffart, S., Marjavaara, S., Paetau, A., Cooper, H., Mattila, K., Lampinen, M., Sajantila, A., Lonnqvist, T., Spelbrink, J. N., and Suomalainen, A. (2008) Infantile-onset spinocerebellar ataxia and mitochondrial recessive ataxia syndrome are associated with neuronal complex I defect and mtDNA depletion, *Human molecular genetics 17*, 3822-3835.

[13] Copeland, W. C. (2014) Defects of mitochondrial DNA replication, *J Child Neurol 29*, 1216-1224.

[14] Yakubovskaya, E., Chen, Z., Carrodeguas, J. A., Kisker, C., and Bogenhagen, D. F. (2006) Functional human mitochondrial DNA polymerase gamma forms a heterotrimer, *J Biol Chem 281*, 374-382.
[15] Euro, L., Farnum, G. A., Palin, E., Suomalainen, A., and Kaguni, L. S. (2011

) Clustering of Alpers disease mutations and catalytic defects in biochemical variants reveal new features of molecular mechanism of the human mitochondrial replicase, Pol {gamma}, *Nucleic acids research*.
[16] Szymanski, M. R., Kuznetsov, V. B., Shumate, C., Meng, Q. C., Lee, Y. S., Patel, G., Patel, S., and Yin, Y. W. (2015) Structural basis for processivity and antiviral drug toxicity in human mitochondrial DNA replicase, *Embo Journal 34*, 1959-1970.
[17] Lee, H. R., and Johnson, K. A. (2006) Fidelity of the human mitochondrial DNA polymerase, *J Biol Chem 281*, 36236-36240.
[18] Wernette, C. M., Conway, M. C., and Kaguni, L. S. (1988) Mitochondrial DNA polymerase from Drosophila melanogaster embryos: kinetics, processivity, and fidelity of DNA polymerization, *Biochemistry 27*, 6046-6054.
[19] Szczepanowska, K., and Foury, F. (2010) A cluster of pathogenic mutations in the 3'-5' exonuclease domain of DNA polymerase gamma defines a novel module coupling DNA synthesis and degradation, *Human molecular genetics 19*, 3516-3529.
[20] Tuske, S., Singh, K., Kaushik, N., and Modak, M. J. (2000) The J-helix of Escherichia coli DNA polymerase I (Klenow fragment) regulates polymerase and 3'- 5'-exonuclease functions, *J Biol Chem 275*, 23759-23768.
[21] Lee, Y. S., Lee, S., Demeler, B., Molineux, I. J., Johnson, K. A., and Yin, Y. W. (2010) Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity, *J Biol Chem 285*, 1490-1499.
[22] Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012) Biomolecular simulation: a computational microscope for molecular biology, *Annual review of biophysics 41*, 429-452.
[23] Miller, B. R., 3rd, Parish, C. A., and Wu, E. Y. (2014) Molecular dynamics study of the opening mechanism for DNA polymerase I, *PLoS computational biology 10*, e1003961.
[24] Lee, Y. S., Kennedy, W. D., and Yin, Y. W. (2009) Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations, *Cell 139*, 312-324.
[25] Brieba, L. G., Eichman, B. F., Kokoska, R. J., Doublie, S., Kunkel, T. A., and Ellenberger, T. (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase, *The EMBO journal 23*, 3452-3461.
[26] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B 102*, 3586-3616.
[27] Donoso, P., Mill, J. G., O'Neill, S. C., and Eisner, D. A. (1992) Fluorescence measurements of cytoplasmic and mitochondrial sodium concentration in rat ventricular myocytes, *J Physiol 448*, 493-509.

[28] Kozoriz, M. G., Church, J., Ozog, M. A., Naus, C. C., and Krebs, C. (2010) Temporary sequestration of potassium by mitochondria in astrocytes, *J Biol Chem 285*, 31107-31119.
[29] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J Comput Chem 26*, 1781-1802.
[30] Mackerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd. (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J Comput Chem 25*, 1400-1415.
[31] Foloppe, N., and MacKerell, A. D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data, *J Comput Chem 21*, 86-104.
[32] MacKerell, A. D., and Banavali, N. K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution, *J Comput Chem 21*, 105-120.
[33] Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *Journal of Chemical Physics 98*, 10089-10092.
[34] Kubo R., T. M., Hashitsume N. (1991) Statistical Physics II Nonequilibrium Statistical Mechanics, *Springer Series 31*.
[35] Feller, S. E., Zhang, Y. H., Pastor, R. W., and Brooks, B. R. (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method, *Journal of Chemical Physics 103*, 4613-4621.
[36] Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant-Pressure Molecular-Dynamics Algorithms, *Journal of Chemical Physics 101*, 4177-4189.
[37] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.
[38] Grossfield, A., and Zuckerman, D. M. (2009) Quantifying uncertainty and sampling quality in biomolecular simulations, *Annu Rep Comput Chem 5*, 23-48.
[39] Luo, N., and Kaguni, L. S. (2005) Mutations in the spacer region of Drosophila mitochondrial DNA polymerase affect DNA binding, processivity, and the balance between Pol and Exo function, *J Biol Chem 280*, 2491-2497.
[40] Stumpf, J. D., Bailey, C. M., Spell, D., Stillwagon, M., Anderson, K. S., and Copeland, W. C. (2010) mip1 containing mutations associated with mitochondrial disease causes mutagenesis and depletion of mtDNA in Saccharomyces cerevisiae, *Human molecular genetics 19*, 2123-2133.
[41] Hakonen, A. H., Isohanni, P., Paetau, A., Herva, R., Suomalainen, A., and Lonnqvist, T. (2007) Recessive Twinkle mutations in early onset encephalopathy with mtDNA depletion, *Brain 130*, 3032-3040.
[42] Baruffini, E., Ferrero, I., and Foury, F. (2007) Mitochondrial DNA defects in Saccharomyces cerevisiae caused by functional interactions between DNA polymerase gamma mutations associated with disease in human, *Biochim Biophys Acta 1772*, 1225-1235.
[43] Wernette, C. M., and Kaguni, L. S. (1986) A mitochondrial DNA polymerase from embryos of Drosophila melanogaster. Purification, subunit structure, and partial characterization, *J Biol Chem 261*, 14764-14770.

[44] Fan, L., Kim, S., Farr, C. L., Schaefer, K. T., Randolph, K. M., Tainer, J. A., and 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, *J Mol Biol 358*, 1229-1243.

Table 1. Model systems and their simulation times. "-" and "+" stand for simulations of the DNA-free and DNA-bound forms of Pol γ, respectively.

Model system	DNA	Pol γ simulation conditions	Number and length of independent simulations	Total simulation time
1	–	WT PDB id: 3IKM (NaCl as salt)	1 x 1 μs	1 μs
2	+	WT PDB id: 3IKM (NaCl as salt)	1 x 1 μs	1 μs
1[a]	–	WT PDB id: 3IKM (KCl as salt)	1 x 400 ns	400 ns
2[a]	+	WT PDB id: 3IKM (KCl as salt)	1 x 400 ns	400 ns
3[b]	+	WT PDB id: 3IKM (with different conformation of the PL in the beginning of each simulation)	1 x 400 ns 1 x 400 ns 1 x 400 ns 1 x 400 ns	1.2 μs
4[b]	+	A467T PDB id: 3IKM	2 x 400 ns	800 ns
5[b]	+	S799A/F800A/W801A PDB id: 3IKM (with different initial structure in the beginning of each simulation)	1 x 400 ns 1 x 400 ns	800 ns
6[b]	+	W748S/E1143G PDB id: 3IKM (MIRAS mutation)	2 x 400 ns	800 ns
7	+	WT (initial structure as in PDB id: 4ZTZ)	1 x 1 μs	1 μs
8	+	WT (initial structure as in PDB id: 4ZTZ, but the first base pair	1 x 1 μs	1 μs

		of bound DNA changed for the canonical one)		

[a] Two independent runs with KCl as salt.

[b] See Methods section for description on how simulation replicas were generated.

Figure 1. Fragments of Pol γ which undergo structural and conformational rearrangements triggered by DNA-binding (left). The schematic diagram of entire Pol γ structure (right). Cyan: the exo domain; green: thumb; yellow: spacer; magenta: the fragments in the catalytic subunit interacting with the distal B subunit (A1 (203-207 aa), A2 (376-383 aa), A3 (524-542) aa); blue: the regions in distal B involved in interaction with the catalytic subunit A [(B1 (246-264 aa), B2 (388-395 aa), B3 (444-451 aa), B4 (458-463 aa)]; orange: the extended partitioning loop (PL; 1044-1095 aa); red: the sequence fragments interacting with DNA [a (499-504 aa), b (583-586 aa), c (755-759 aa), d (768-770 aa), e (802-807 aa), f (858-861 aa) and the extended orienter (304-316 aa)]. The flexible parts in the spacer are highlighted with purple (618-673 aa) and dark blue (723-750 aa), respectively. Labeling is based on the order of fragments in the sequence. Assigned entire sequence of the catalytic subunit is shown (lower panel). All previously studied fragments which fall into the areas of conformational changes discussed in the current work are marked with preserved original assignment at the bottom of the sequence panel. For clarity, proximal B subunit and NTD domain as well as major part of the pol domain in the catalytic A subunit have been omitted from the figure. PDB id: 3IKM [24].

Figure 2. Conformational changes in the Pol γ structure revealed by atomistic simulations. The upper panel represents different snapshots from the 1 μs simulation of DNA-bound Pol γ (model system 2 in Table 1), ranging from the starting structure to the final configuration at 1 microsecond.

The lower panel represents similar snapshots based on the 1 μs simulation of DNA-free Pol γ (model system 1 in Table 1). The snapshots chosen for the figure highlight distinct structural changes that occur during the MD simulations. Color code is as in Figure 1. DNA (upper panel) is shown in surface representation.

Figure 3. Structural rearrangements in the S799/F800/W801 regions in WT and triple Ala mutant upon loading on DNA. The upper and lower panels display snapshots from the start and the end of the simulation. DNA (white) is shown in tube representation in panels B and C.

Figure 4. Perturbation in DNA-protein binding as a consequence of the triple alanine mutation of the SFW motif (model system # 5). The distance is plotted between the P atom of the nearest nucleic acid and the Cα atom of the amino acid. The results from both simulation replicas 1 (A) and 2 (B) are shown.

Figure 5. Dynamics of the partitioning loop (PL) and orienter upon DNA binding. A) Mutual position of the orienter segment and the PL in apo Pol γ in the end of the 1 μs simulation (system 1 in Table 1). B) In DNA-bound Pol γ (system 2 in Table 2) the PL together with the orienter segment complete the DNA-binding channel in the pol mode of the enzyme. Cyan: residues whose mutations result in high mutation rates and the exo phenotype. Inset of panel B: interactions between these residues and the template DNA strand. Green and gray: hydrogen bonds; magenta: hydrophobic interactions. DNA (white) is shown in a surface representation.

Figure 6. Rotation of the exonuclease subdomain. The distance is plotted between the conserved and buried R227 (Cα atom) and the nearest DNA backbone atom (the P atom of nucleic base 1012) for model systems 2, 7 and 8. The horizontal line shows the distance between the same residue and the DNA in DNA-bound crystal structure (PDB id: 4ZTZ).

Figure 7. Conformational changes in the spacer in the region of Cluster 5 upon DNA binding. Simulation of A) WT enzyme without bound DNA (model system 1), B) WT enzyme with bound DNA (model system 2), C) W748S mutant with bound DNA (model system 6), and D) A467T mutant with bound DNA (model system 4). All structures represent the final configuration of the given simulations. Magenta: carbon atoms of residues from Cluster 5. Cyan: carbon atoms of residues from the GWY (GWF in fly) motif. DNA is not shown for clarity.

Figure 8. Distance between Cα of His733 and P of 1017 (DNA base) as a function of simulation time in two different simulations of DNA-bound form (model system 2 (A) and 2^a (B)). The black horizontal line shows the same distance in the crystal structure of the apo enzyme.

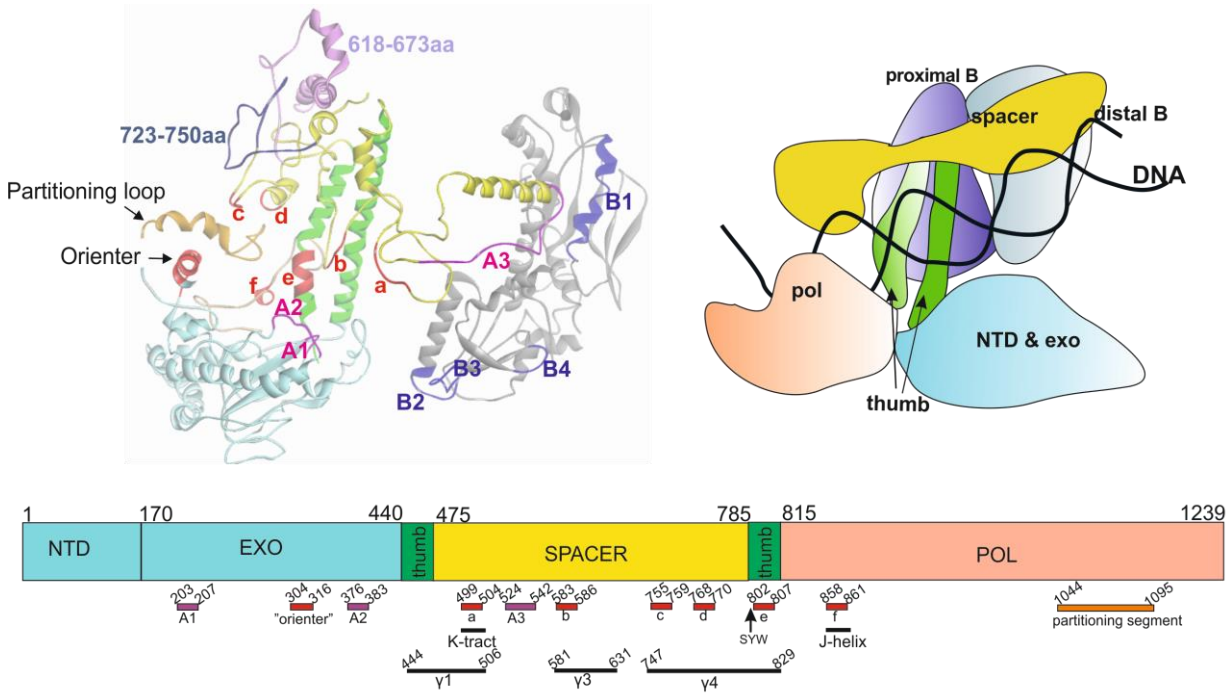

618-673aa
723-750aa
Partitioning loop
Orienter
c
d
b
e
f
a
A1
A2
A3
B1
B2
B3
B4
proximal B
spacer
distal B
DNA
pol
NTD & exo
thumb
1
170
440
475
785
815
1239
NTD
EXO
thumb
SPACER
thumb
POL
203
207
A1
304
316
"orienter"
376
383
A2
499
504
a
524
542
A3
583
586
b
755
759
c
768
770
d
802
807
e
SYW
858
861
f
J-helix
K-tract
444
506
γ1
581
631
γ3
747
829
γ4
1044
1095
partitioning segment

Figure 1

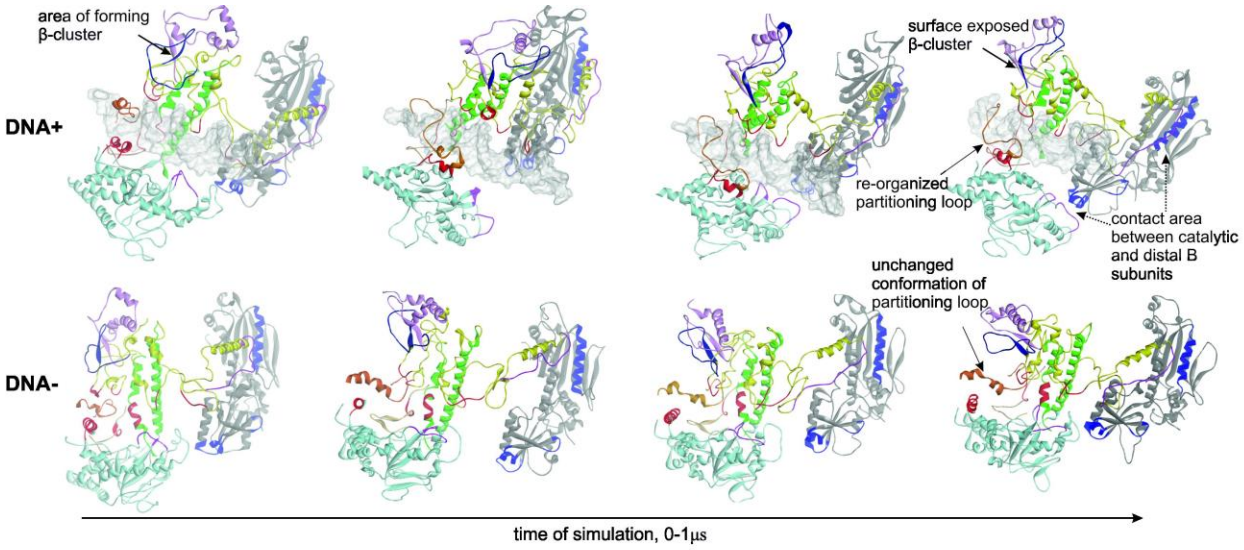

area of forming
β-cluster
surface exposed
β-cluster
DNA+
re-organized
partitioning loop
contact area
between catalytic
and distal B
subunits
unchanged
conformation of
partitioning loop
DNA-
time of simulation, 0-1μs

Figure 2

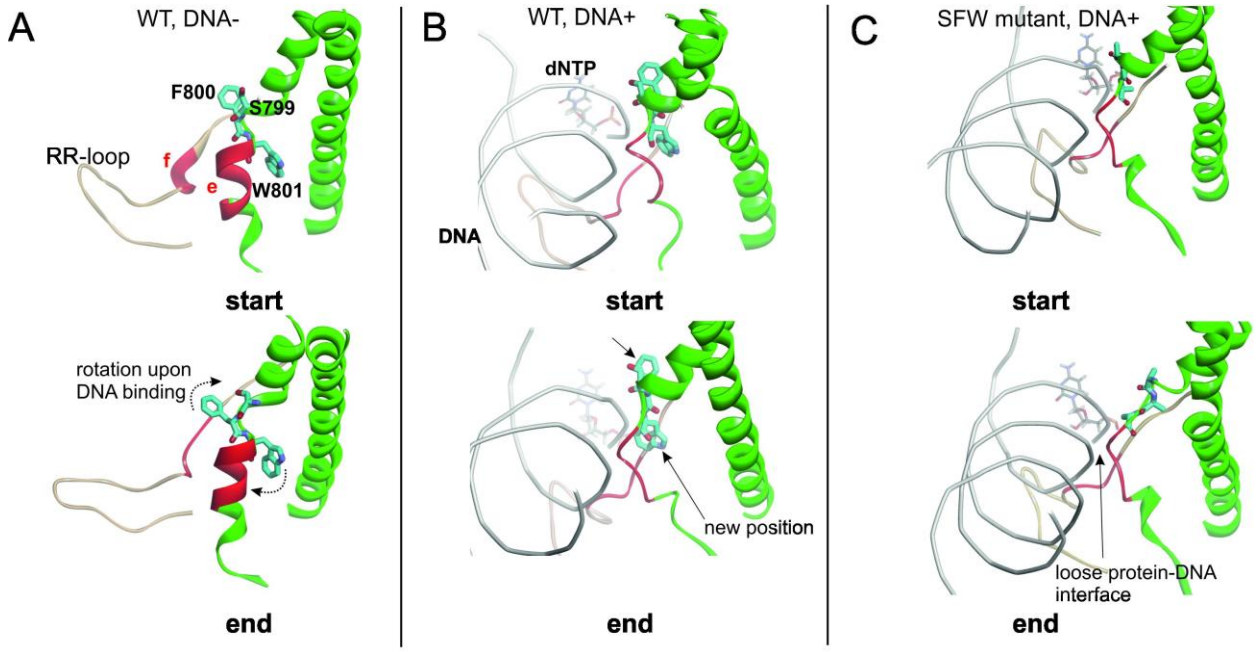

A
WT, DNA-
F800
S799
RR-loop
f
e
W801
start
rotation upon
DNA binding
end
B
WT, DNA+
dNTP
DNA
start
new position
end
C
SFW mutant, DNA+
start
loose protein-DNA
interface
end

Figure 3

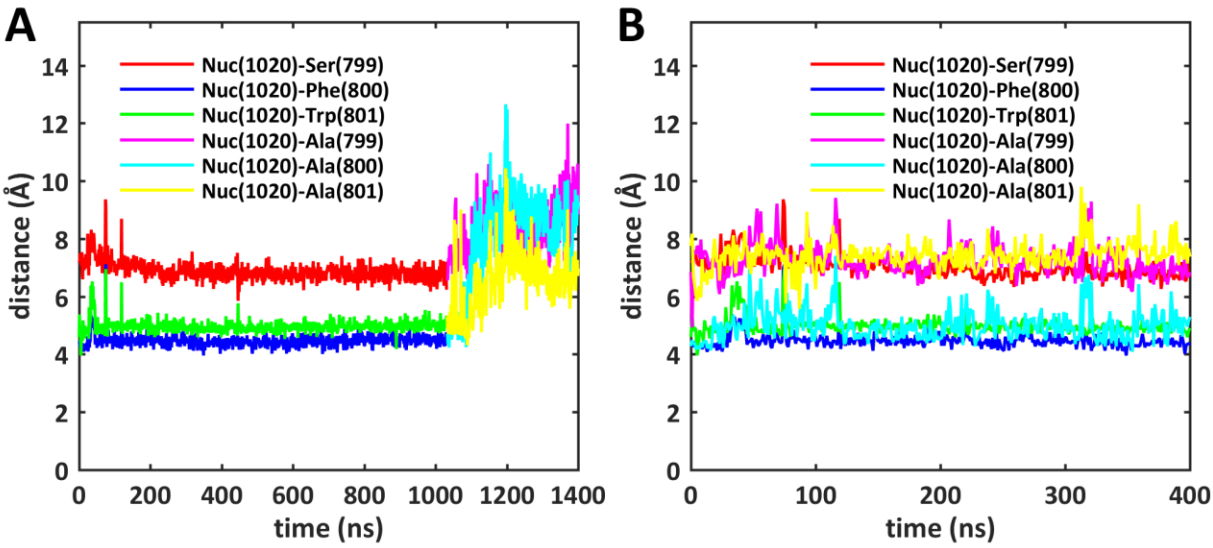

A
B
Nuc(1020)-Ser(799)
Nuc(1020)-Phe(800)
Nuc(1020)-Trp(801)
Nuc(1020)-Ala(799)
Nuc(1020)-Ala(800)
Nuc(1020)-Ala(801)
distance (Å)
time (ns)
0
2
4
6
8
10
12
14
0
200
400
600
800
1000
1200
1400
100
300

Figure 4.

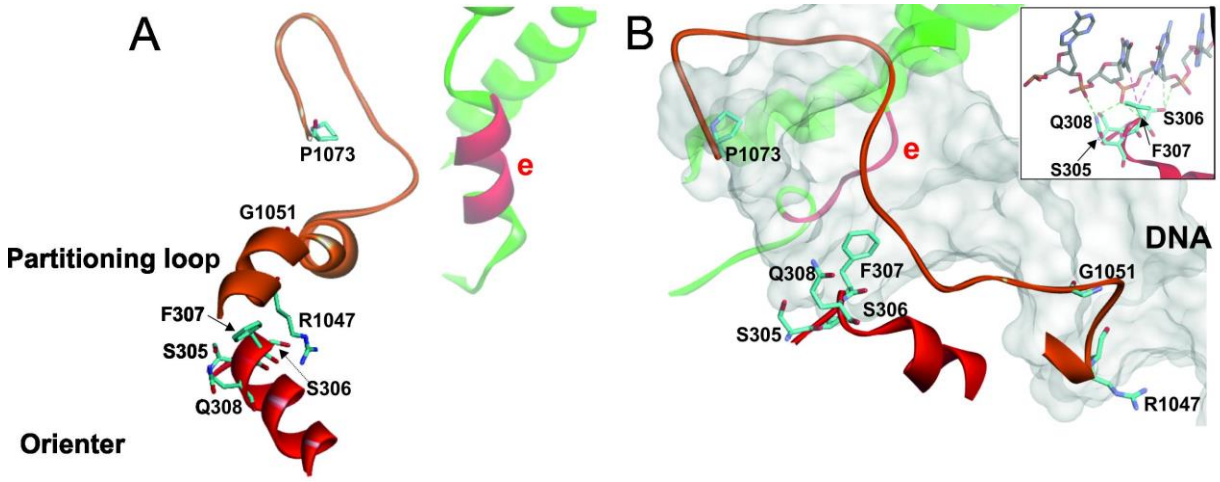

A
P1073
e
G1051
Partitioning loop
F307
R1047
S305
S306
Q308
Orienter
B
P1073
e
Q308
S306
F307
S305
DNA
Q308
F307
G1051
S306
S305
R1047

Figure 5

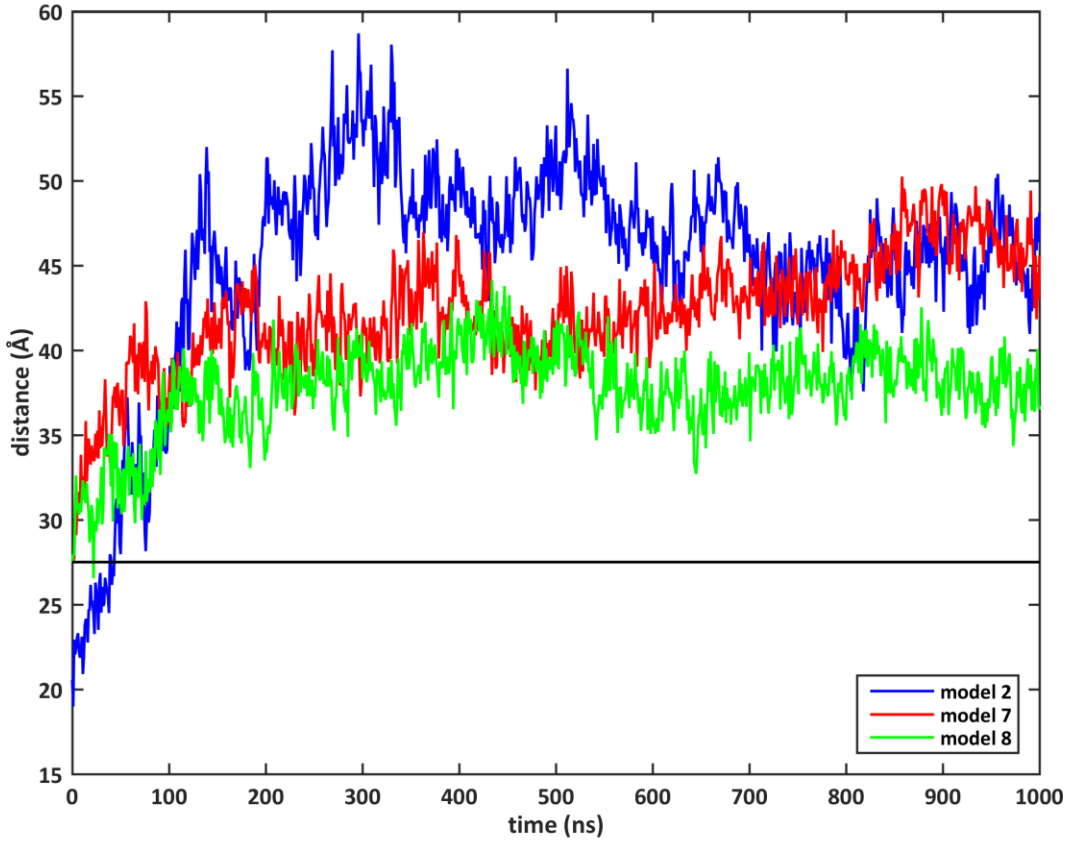

distance (Å)
time (ns)
model 2
model 7
model 8
60
55
50
45
40
35
30
25
20
15
0
100
200
300
400
500
600
700
800
900
1000

Figure 6.

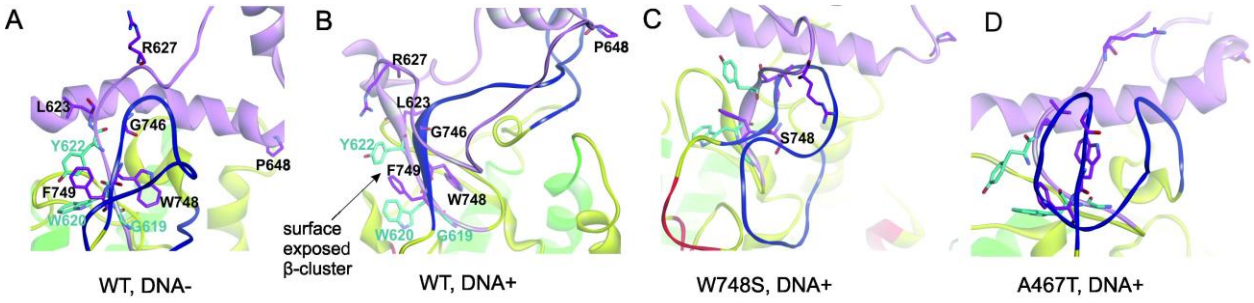

A
R627
L623
G746
Y622
P648
F749
W748
W620
G619
WT, DNA-
B
R627
P648
L623
G746
Y622
F749
W748
surface
exposed
β-cluster
W620
G619
WT, DNA+
C
S748
W748S, DNA+
D
A467T, DNA+

Figure 7

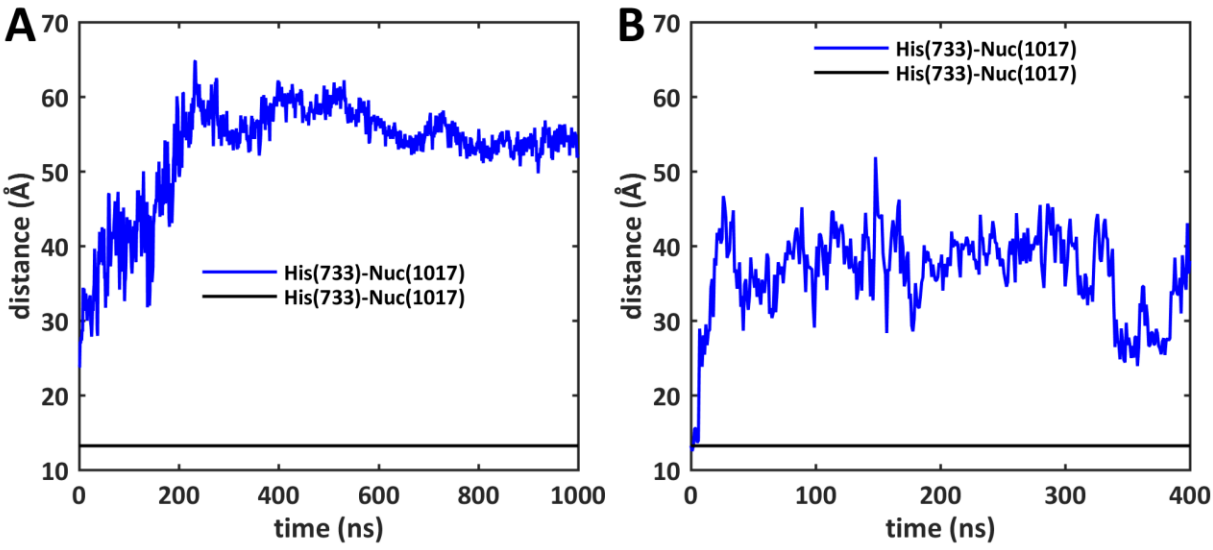
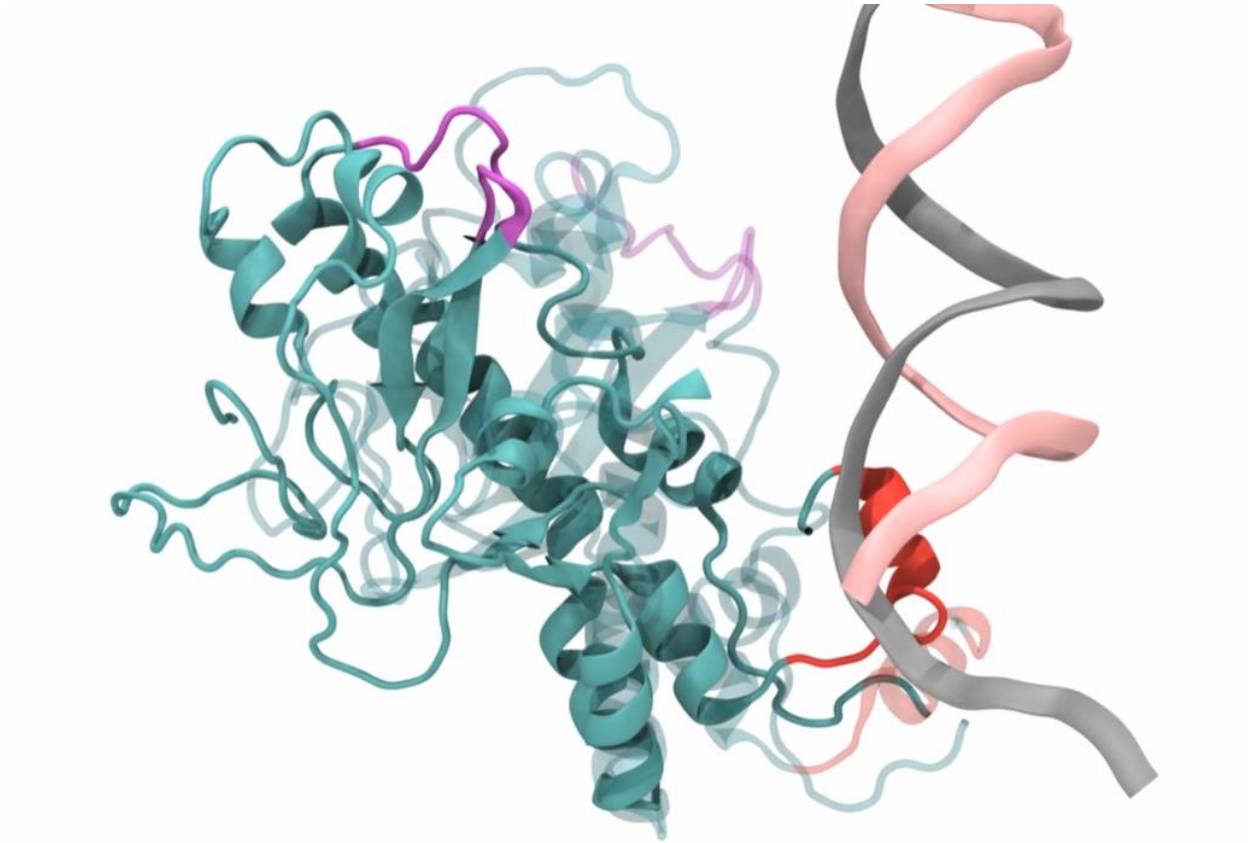


Figure 8

For Table of Contents use only



Atomistic Molecular Dynamics Simulations of Mitochondrial DNA Polymerase Gamma: Novel Mechanisms of Function and Pathogenesis

Liliya Euro, Outi Haapanen, Tomasz Rog, Ilpo Vattulainen, Anu Suomalainen, Vivek Sharma