Functional and structural properties of eukaryotic DNA polymerase epsilon

Olga Chilkova

Department of Medical Biochemistry and Biophysics
Umeå University, Sweden
Umeå 2006
Моим маме и папе
# Table of contents

ABSTRACT ........................................................................................................... 7

PAPERS INCLUDED IN THE THESIS ................................................................. 9

PAPERS NOT INCLUDED IN THE THESIS ......................................................... 10

1. INTRODUCTION ............................................................................................ 11
   Assembly of the eukaryotic replication fork ................................................. 12
   Replicative DNA polymerases .................................................................... 12
     Pol α ........................................................................................................ 13
     Pol δ ........................................................................................................ 14
     Pol ε ........................................................................................................ 14
   SV40 model of eukaryotic DNA replication .............................................. 17
   Processivity ................................................................................................. 17
     Replication protein A ............................................................................. 17
     Proliferating cell nuclear antigen and replication factor C ................. 18
   Fidelity ....................................................................................................... 18

2. AIMS ............................................................................................................. 20

3. RESULTS ....................................................................................................... 21
   The quaternary structure of DNA polymerase epsilon from Saccharomyces cerevisiae (Paper I) .............................................................. 21
   Structure of DNA polymerase epsilon by cryo electron microscopy (Paper II) .................................................................................. 22
   Unique error signature of the four-subunit yeast DNA polymerase epsilon (Paper III) ................................................................. 22
   Similarities and differences between the two replicative DNA polymerases, DNA polymerase δ and DNA polymerase ε (Paper IV)  24

4. CONCLUSIONS ............................................................................................. 26

ACKNOWLEDGMENTS ....................................................................................... 27

REFERENCES .................................................................................................... 28
Abstract

Functional and structural properties of eukaryotic DNA polymerase epsilon

In eukaryotes there are three DNA polymerases which are essential for the replication of chromosomal DNA: DNA polymerase α (Pol α), DNA polymerase δ (Pol δ) and DNA polymerase ε (Pol ε). In vitro studies of viral DNA replication showed that Pol α and Pol δ are sufficient for DNA replication on both leading and lagging DNA strands, thus leaving the function of Pol ε unknown. The low abundance and the reported protease sensitivity of Pol ε were holding back biochemical studies of the enzyme. The aim of this study was to characterize the structural and functional properties of eukaryotic Pol ε.

We first developed a protocol for over-expression and purification of Pol ε from the yeast Saccharomyces cerevisiae. Pol ε consists of four subunits: Pol2 (catalytic subunit), Dpb2, Dpb3 and Dpb4. This four-subunit complex was purified to homogeneity by conventional chromatography and the subunit stoichiometry of purified Pol ε was estimated from colloidal coomassie-stained gels to be 1:1:1:1. The quaternary structure was determined by sedimentation velocity and gel filtration experiments. Molecular mass (371 kDa) was calculated from the experimentally determined Stokes radius (74.5 Å) and sedimentation coefficient (11.9 S) and was in good agreement with a theoretical molecular mass calculated for a heterotetramer (379 kDa). Analytical sedimentation equilibrium ultracentrifugation experiments supported the proposed heterotetrameric structure of Pol ε.

By cryo-electron microscopy and single-particle image analysis we determined the structure of Saccharomyces cerevisiae Pol ε to 20-Å resolution. The four-subunit complex was found to consist of a globular domain, comprising the Pol2 subunit, flexibly connected to an elongated domain, including Dpb2, Dpb3 and Dpb4 subunits. We found that Pol ε requires a minimal length of 40 base pairs of primer-template duplex to be processive. This length corresponds to the dimensions of the elongated domain.

To characterize the fidelity by which Pol ε synthesizes DNA, we purified wild type and exonuclease-deficient Pol ε. Wild type Pol ε synthesizes DNA with a very high accuracy. Analysis of the exonuclease-deficient Pol ε showed that Pol ε proofreads more than 90% of the errors made by its polymerase activity. Exonuclease-deficient Pol ε was shown to have a specific spectrum of errors not seen in other DNA polymerases: a high proportion of transversions resulting from T-dTTP, T-dCTP and C-dTTP mispairs. This unique error specificity and amino acid sequence alignment suggest that the structure of the polymerase active site of Pol ε differs from those of other members of B family DNA polymerases.

With recombinant proteins and circular single-stranded DNA templates, we partially reconstituted DNA replication in vitro, in which we challenged Pol ε and Pol δ in side-by-side comparisons regarding functional assays for polymerase activity and processivity, as well as physical interactions with nucleic acids and PCNA. We found that Pol ε activity and “on-DNA” PCNA interactions are dependent on RPA-coated template DNA. By the surface plasmon resonance technique, we showed that Pol ε has a high affinity for DNA and low affinity for immobilized PCNA. By contrast, Pol δ was found to have low affinity for DNA and high affinity for PCNA. We suggest that a possible function of RPA is to regulate down the DNA synthesis through Pol ε, and that the mechanism by which Pol ε and Pol δ load onto the template is different due to different properties of the interaction with DNA and PCNA.
Papers included in the thesis

This thesis is based on the following papers, which are referred to by Roman numerals (I-IV).

I. The quaternary structure of DNA polymerase epsilon from *Saccharomyces cerevisiae*.
   Chilkova O., Jonsson B.H., Johansson E.

II. Structure of DNA polymerase epsilon by cryo electron microscopy.
   Asturias F.J., Cheung I.K., Sabouri N., Chilkova O., Wepplo D., Johansson E.

III. Unique error signature of the four-subunit yeast DNA polymerase epsilon.
   Shcherbakova P.V., Pavlov Y.I., Chilkova O., Rogozin I.B., Johansson E., Kunkel T.A.

IV. Similarities and differences between the two replicative DNA polymerases, DNA polymerase δ and DNA polymerase ε.
   Chilkova O., Stenlund P., Isoz I., Grabowski P., Lundström E.-B., Johansson E.
   Manuscript
Papers not included in the thesis

Contributions to the following publications, not included in this thesis, were also made:

1. **Enzymatic switching for efficient and accurate translesion DNA replication.**
   
   McCulloch S.D., Kokoska R.J., Chilkova O., Welch C.M., Johansson E., Burgers P.M.J., Kunkel T.A.
   

2. **The *Schizosaccharomyces pombe* replication inhibitor Spd1 regulates ribonucleotide reductase activity and dNTPs by binding to the large Cdc22 subunit.**

   Håkansson P., Dahl L., Chilkova O., Domkin V., Thelander L.
   
1. Introduction

Genetic information in all organisms, as well as in many viruses, is stored in the form of DNA. Reproduction of genetic information by DNA replication is one of the most important processes in a living cell. DNA has to be replicated quickly and precisely to maintain genetic information without errors. Such errors can lead to mutations, which may cause cell death, inheritable diseases or cancer. Eukaryotes utilize complex multiprotein replication machineries (Fig. 1) for the replication of their large genomes.

![Fig.1. Hypothetical placement of protein complexes at the eukaryotic replication fork.](image)

The enzyme responsible for the DNA synthesis is always a DNA polymerase (Pol). Three DNA polymerases (Pols) are essential for the bulk synthesis of DNA during chromosomal DNA replication in eukaryotic cells: DNA polymerase α (Pol α), DNA polymerase δ (Pol δ) and DNA polymerase ε (Pol ε).

This thesis describes functional and structural properties of Pol ε. In the following section the current views on the eukaryotic replication machinery and replicative Pols are reviewed.
Assembly of the eukaryotic replication fork

DNA replication takes place in S phase of the cell cycle. The preparation to this process starts early in the G1 phase of the cell cycle by formation of a pre-replication complex (pre-RC) consisting of an origin recognition complex (ORC), Cdc6, Cdt1 and Mcm2-7 [reviewed in 1,2]. ORC is a six-subunit complex, highly conserved in eukaryotes. In yeast ORC is bound to the origins of replication throughout the cell cycle [reviewed in 3]. ORC sequentially recruits Cdc6 and Cdt1 and utilizes ATP hydrolysis to load the six-subunit ring-shaped Mcm2-7 complex [4]. Mcm2-7 is believed to be a eukaryotic helicase, which unwinds double-stranded DNA (dsDNA) [reviewed in 5].

Pre-RC turns to a post-RC in G1/S transition via activation by S-phase cyclin-dependent kinases and Cdc7-Dbf4 kinase [reviewed in 6]. For the initiation of replication a number of factors bind to the origin simultaneously and interdependently [reviewed in 2,7]. Mcm10 is recruited to the replication origin and facilitates loading of Cdc45 [8]. In Xenopus Cdc45 was shown to stimulate helicase activity of the Mcm2-7 complex [9]. Mcm10 forms a complex with Pol α-primase and recruits it to the replication origin [10]. At the same time, replication protein A (RPA) binds the origin [11]. Initiation factor Sld2 stimulates recruitment of Dpb11 to the replication origin [12]. The Dpb11/Sld2 complex and a four-subunit GINS complex have been suggested to load Pol ε onto the origin [13,14]. Replication factor C (RFC) loads proliferating cell nuclear antigen (PCNA), which facilitates loading of Pol δ, and DNA synthesis starts.

Replicative DNA polymerases

There are 16 Pols at the eukaryotic cell [15]. They are classified into four families on the basis of their primary structure [16]. Pol α, Pol δ and Pol ε are included in the B family of Pols by homology with the E. coli polB gene product, Pol II (Table 1). The B family of Pols also includes eukaryotic Pol ζ, several archaeal Pols and replicative Pols of bacteriophages T4 and RB69. There are no structures available for multi-subunit eukaryotic B Pols. However, crystal structures of single-subunit B Pols from bacteriophage RB69 [17] and from several archaeal species [18-21] have been solved. All Pols have common structural domains termed “thumb”, “palm” and “fingers” with a catalytic site situated at the “palm” domain [reviewed in 22].

* Unless specifically indicated, all discussed processes and factors are found in yeast Saccharomyces cerevisiae.
Table 1. Eukaryotic replicative DNA polymerases

<table>
<thead>
<tr>
<th>Name</th>
<th>Subunit composition in S. cerevisiae (kDa)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Pol1 (180)</td>
<td>Priming in DNA replication</td>
</tr>
<tr>
<td></td>
<td>Pol12 (86)</td>
<td>S phase checkpoint control</td>
</tr>
<tr>
<td></td>
<td>Pri1 (48)</td>
<td>Lagging strand DNA polymerase</td>
</tr>
<tr>
<td></td>
<td>Pri2 (58)</td>
<td>DNA repair and recombination</td>
</tr>
<tr>
<td>δ</td>
<td>Pol3 (125)</td>
<td>DNA replication, repair and recombination</td>
</tr>
<tr>
<td></td>
<td>Pol31 (50)</td>
<td>S phase checkpoint control</td>
</tr>
<tr>
<td></td>
<td>Pol32 (40)</td>
<td></td>
</tr>
<tr>
<td>ε</td>
<td>Pol2 (256)</td>
<td>DNA replication, repair and recombination</td>
</tr>
<tr>
<td></td>
<td>Dpb2 (78)</td>
<td>S phase checkpoint control</td>
</tr>
<tr>
<td></td>
<td>Dpb3 (23)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dpb4 (22)</td>
<td></td>
</tr>
</tbody>
</table>

Pol α

Pol α possesses a unique ability to initiate a new DNA chain by synthesis of a RNA-DNA hybrid primer on single-stranded DNA (ssDNA). Yeast Pol α consists of four subunits, Pol1 (180 kDa), Pol12 (86 kDa), Pri2 (58 kDa) and Pri1 (48 kDa) encoded by essential genes [23]. This four-subunit structure is highly conserved among eukaryotes [24]. The largest Pol1 subunit is a DNA polymerase. Pri1 and Pri2 comprise primase. Pol12, or B subunit, is important for cell cycle regulation of replication initiation by Pol α [25]. After recruitment to the initiation complex at the replication origin, the primase subunit synthesizes a short ~10 nucleotides (nt) long RNA primer, which is then elongated by a stretch of 20-30 nt by the Pol1 subunit of Pol α [26]. In the same manner, Pol α synthesizes primers for each Okazaki fragment on the lagging strand of the replication fork. It is still unclear what mechanisms regulate priming by Pol α at every ~200 nt, the average length of an Okazaki fragment. The primer, which is synthesized by Pol α, is elongated further by one of the replicative Pols: Pol δ or Pol ε.
Pol δ

Pol δ from *S. cerevisiae* consists of three subunits, A/Pol3 (125 kDa), B/Pol31 (50 kDa) and C/Pol32 (40 kDa) [27]. Subunit A is a highly conserved essential catalytic subunit, with 3'-5' proofreading activity [28]. Subunit B forms a stable complex with subunit A, is highly conserved and is essential for viability [29-31]. Subunit C is not essential for viability, but its deletion mutants show replication and repair defects [29]. The carboxy-terminus of subunit C contains a PCNA-binding motif, important for interaction of Pol δ with PCNA in solution [32]. In *S. pombe* and humans Pol δ contains a fourth subunit, D [33,34]. Subunit D stabilizes the complex of subunits A and B and participates in interaction with PCNA [35-37]. Originally recombinant Pol δ was proposed to form a dimer of heterotrimers with a molecular mass over 500 kDa [38]. The quaternary structure of Pol δ was refined later, using combined gel filtration and glycerol gradient analysis. Pol δ was shown to be a heterotrimer with each subunit, Pol3, Pol31 and Pol32, represented once [39]. The elongated shape of Pol32 caused the overestimation of the molecular mass in earlier gel filtration experiments [32,40].

Several recent studies have assigned Pol δ to the lagging strand DNA synthesis [reviewed in 7]. After the initiation of an Okazaki fragment by Pol α, RFC-PCNA complex terminates DNA synthesis by Pol α, and Pol δ elongates the rest of the Okazaki fragment [41-45]. Pol δ has also been shown to correct errors made by Pol α *in vivo* [46]. When Pol δ meets the 5' end of the previously synthesized Okazaki fragment it displaces the downstream primer [47]. The displaced short flap is degraded by flap endonuclease 1 (FEN1), and the resulting nick is ligated by DNA ligase I [48]. In the absence of Okazaki fragment maturation factors FEN1 and DNA ligase I, Pol δ is able to degrade newly replicated DNA using its 3'-5' exonuclease activity. The degradation is again followed by resynthesis and displacement until the flap is degraded by FEN1. This process is called "idling" [49]. Due to idling activity, Pol δ maintains a ligatable nick without significant displacement of the 5' end of the previously synthesized Okazaki fragment. In the same study Pol ε was shown to lack this idling activity, leaving Pol δ to be solely responsible for the lagging strand DNA synthesis [49].

Pol ε

Pol ε has been the most enigmatic Pol since its precise function in eukaryotic cells is unclear despite numerous studies [reviewed in 7,50,51]. Yeast Pol ε consists of four subunits, Pol2 (256 kDa), Dpb2 (78 kDa), Dpb3 (23 kDa) and Dpb4 (22 kDa) [52].
Pol2 is a catalytic subunit with 3′-5′ exonuclease activity. The structure of Pol2 differs from the catalytic subunits of other B family Pols by the presence of a large carboxy-terminal domain (Fig. 2). Conserved motifs in the catalytic center of Pol2 are different from those of other B Pols mainly in residues involved in template-primer binding [53]. Pol2 has been shown to be essential for viability since cells with a deleted POL2 gene and temperature-sensitive mutations in the catalytic domain fail to replicate chromosomal DNA [54-57]. However, the amino-terminal catalytic domain could be deleted without loss of viability, although this deletion causes defects in growth and in cell cycle progression of DNA replication [58-61]. The carboxy-terminal domain is essential for growth and, probably, participates in the initiation of DNA replication and in S/M checkpoint control [59,62,63]. The putative zinc finger region of the C-terminal domain is important for the interaction with Dpb2 [63,64]. Dpb2 is a conserved essential for viability subunit which function is unknown [65-68]. In fission yeast Dpb2 is essential for viability since it binds DNA replication origins early in S phase [68]. Cyclin-dependent kinase Cdc28 phosphorylates Dpb2 in multiple sites and this phosphorylation is suggested to be important for the interaction with Pol2 and for the activity of Pol ε complex [69]. The two small Pol ε subunits Dpb3 and Dpb4 are not essential for viability in S. cerevisiae, however, Dpb3 is required for cell cycle progression in S. pombe [70-72]. Homologues of Dpb3 are found in human, mouse and Drosophila [72,73]. Dpb3 forms a complex with Dpb4 and interacts with Dpb2 [72,73]. Dpb4 has been reported to play an important role in maintaining the subunit structure of Pol ε [71]. Both the Dpb3 and the Dpb4 subunits possess histone-fold motifs, whereas Dpb4 was found to be part of a chromatin remodeling complex [71,73-75].

Protease sensitivity and low levels of the enzyme in cells thwarted purification of Pol ε [52,76]. Previously, recombinant Pol ε was shown to be a dimer in vitro, using gel filtration studies similar to early studies on Pol δ [38,63]. Due to limited amounts, subunit stoichiometry of Pol ε was estimated from silver stained gels and was reported to be 1:1:3:4 (Pol2:Dpb2:Dpb3:Dpb4) [52,77]. Functional in vitro characterization of Pol ε is limited. Yeast and human Pol ε has been shown to possess high intrinsic processivity on homopolymer poly(dA) - oligo(dT) synthetic template and to be stimulated by PCNA [52,78,79]. Recently, a novel clamp-like complex GINS was shown to stimulate the rate of DNA synthesis by Pol ε [80]. Pol ε
has been reported to interact with single-stranded and double-stranded DNA in gel-shift assays [81,82].

Over the years, Pol ε has been proposed to play a role in several cellular processes. According to the genetic studies mentioned above Pol ε participates in the assembly of an initiation complex early in S phase. This assignment is supported by the findings that Pol ε loads to replication origins at the same time as Pol α, i.e. before priming of DNA synthesis [13,83,84]. Pol ε has been suggested to replicate silent regions of chromatin and to participate in the maintenance of chromatin complexes at boundaries between active and silent chromatin [75]. Pol ε was also shown to participate in transcriptional silencing in yeast [85]. A role in sister chromatid cohesion was assigned to Pol ε because of the genetic and physical interaction between Pol2 and genes TRF4 and TRF5. This interaction was suggested to be a coordinating link between DNA replication and sister chromatid cohesion [86,87]. Pol ε has been suggested to replicate specific parts of the genome DNA, according to studies by confocal laser-scanning microscopy and immuno electron microscopy in human cells, where Pol ε and Pol δ have been shown to be separated spatially and temporally in early and late replicated chromatin [88,89].

There are several lines of evidence that Pol ε replicates chromosomal DNA in cells, although it is not essential for SV40 DNA replication [90]. The above-mentioned in vivo mutation studies of Pol2 catalytic subunit indicate that Pol ε participates in cellular DNA replication. Pol ε has been UV-crosslinked to cellular chromatin in monkey cells in the same manner as Pol α and Pol δ [91]. Using chromatin immunoprecipitation (ChIP) it was shown that Pol ε moved together with Mcm and Cdc45 as the replication fork progresses [83]. With similar ChIP and immunohistochemical assays Pol ε colocalized with Pol α and Pol δ and replication origins early in S phase [92]. DNA synthesis was significantly inhibited in nuclei of human cells injected with neutralizing antibodies to Pol2 subunit [93]. DNA replication is significantly impaired in Xenopus egg extracts, immunodepleted with antibody raised against the second largest subunit of Pol ε [94,95]. Proofreading deficient mutants of Pol ε have increased mutation rate, which indicates that Pol ε participates in bulk of DNA synthesis [96,97].

The studies mentioned here led to a hypothesis that one of the two Pols, Pol ε or Pol δ, performs the bulk of DNA synthesis on the leading strand and the other on the lagging strand. This hypothesis is still open for discussion and is one of the most contradictory in the literature [7,98]. Studies showing strand specificity of error correction in chromosomal and plasmid DNA in proofreading deficient Pol ε and Pol δ mutants [99,100] support the “one strand – one DNA polymerase” hypothesis. In addition, an asymmetry in replication errors on opposite strands is established at the replication forks [101].
**SV40 model of eukaryotic DNA replication**

Initially studies on eukaryotic DNA replication were carried out on cell-free extracts using plasmid DNA, containing a simian virus 40 (SV40) origin [102-104]. Reconstitution of SV40 replication in vitro determined a minimal set of proteins required for the complete replication of the viral genome [105,106]: large tumor SV40 antigen (origin recognition and helicase activities), Pol α, Pol δ, RPA, RF-C, PCNA, FEN1, DNA ligase I and DNA topoisomerases I an II. For mechanistic in vitro studies of replicative Pols partial reconstitution of DNA replication with recombinant factors is widely used.

**Processivity**

Pols that replicate chromosomal DNA have to be fast and precise in order to replicate the entire genome DNA in a reasonable time and with the highest accuracy possible.

The rate of DNA synthesis depends on such processes as primer end detection and loading onto primer end, as well as on the ability to synthesize a long DNA stretch without dissociation from the primer end. The variable characterizing such ability is called processivity [107]. For the measurement of Pols processivity synthetic (poly(dA) – oligo(dT)) or natural (single-stranded single-primed circular DNA) templates are used. Synthesized products are analyzed using either internal labeling with a radioactive dNTP added to the reaction or an end labeled primer annealed to a single-stranded template. An essential criterion for proper processivity measurement is a single-hit condition [107]. It ensures that a polymerase molecule elongates only a primer which was not extended previously by another polymerase molecule. To achieve this, the following approaches are used: 1) a high ratio between primer-template and enzyme; 2) using a trap molecule, which prevents the Pol from binding a new substrate after dissociation from the extended primer. Single-hit condition is achieved when a) less than 20% of the primer is utilized in the reaction [108], b) termination probability at certain nucleotide is equal at various incubation time points [109].

The processivity of a Pol depends on several factors, among them are: the sequence and structure of the DNA primer-template and the contribution of specific processivity factors.

**Replication protein A**

Secondary structures in a single-stranded template, such as hairpins, might cause Pols to dissociate from the primer-template. Also unproductive binding to ssDNA traps Pol ε from the template-primer junction [81,110].
The eukaryotic ssDNA binding protein, RPA, stabilizes DNA in a single-stranded conformation [reviewed in 111,112]. RPA is a conserved heterotrimer composed of RPA1 (70 kDa), RPA2 (30 kDa) and RPA3 (14 kDa) subunits which are encoded by genes essential for viability [113]. High ssDNA affinity of about $10^9 M^{-1}$ is a characteristic property of RPA [114]. RPA was first characterized as an essential component of SV40 DNA replication \textit{in vitro} [115]. In addition to DNA replication, RPA plays a significant role in DNA repair and recombination [reviewed in 116].

**Proliferating cell nuclear antigen and replication factor C**

Among nuclear Pols Pol $\varepsilon$ and Pol $\delta$ are the most processive. Pol $\varepsilon$ is intrinsically processive, but Pol $\delta$ achieves high processivity with the help of the processivity clamp PCNA and clamp loader RFC [38,52]. The yeast \textit{POL30} gene, encoding PCNA, is essential for viability [117]. PCNA is a ring-like homotrimer of a 29 kDa polypeptide which is able to encircle double-stranded DNA and slide freely along it [118]. PCNA is loaded onto the template-primer junction by the clamp loader RFC. RFC consists of five conserved subunits: Rfc1 (95 kDa), Rfc2 (40 kDa), Rfc3 (38 kDa), Rfc4 (36 kDa) and Rfc5 (40 kDa). Loading of PCNA by RFC is a multistep, ATP dependent process [44,45]. Besides DNA replication, PCNA participates in DNA repair, DNA recombination and cell cycle control via its interactions with various proteins involved in these processes [reviewed in 119].

Pol $\delta$ interacts with PCNA in two distinct modes “off-DNA” and “on-DNA” [reviewed in 7]. For interaction in solution, Pol $\delta$ depends on a consensus PCNA-binding motif QxxLxxFF common for many PCNA-binding proteins, located in the carboxy-terminal part of the Pol32 subunit [32,120]. However, the Pol3 subunit interacts with PCNA, when PCNA is bound onto DNA [reviewed in 7]. DNA synthesis by Pol $\varepsilon$ is stimulated by PCNA \textit{in vitro} [78]. The Pol2 subunit of Pol $\varepsilon$ contains a consensus PCNA-binding motif; however, its importance for interaction with PCNA \textit{in vitro} has not been studied. Although Pol $\varepsilon$ is stimulated by PCNA \textit{in vitro} [52,78,79], site-directed mutagenesis studies of this conserved motif showed no lethality. However, mutant cells were sensitive to a DNA-alkylating agent, methyl methanesulfonate [77].

**Fidelity**

Accuracy or fidelity of DNA synthesis during replication of chromosomal DNA is crucial to maintain genomic stability. Pol $\varepsilon$ and Pol $\delta$ are the only nuclear Pols that possess intrinsic 3'-5' exonuclease activity. Mutations that inactivate this proofreading activity of either Pol $\varepsilon$ or Pol $\delta$ lead to increased mutation rates in yeast and to cancer in mice [96,121]. Mutation rates in Pol
δ proofreading deficient yeast cells is about 10 times higher than in Pol ε proofreading deficient yeast cells [96]. The fidelity of a purified Pol can be measured using one of two methods, either a primer-extension assay or a gap-filling assay on a M13mp2 template. The primer-extension assay is a gel-based method where the kinetics of correct or incorrect nucleotide incorporation is measured using a synthetic oligonucleotide primer-template [108]. In this method products from a polymerase reaction are resolved by polyacrylamide gel electrophoresis and quantified with a phosphoimager. Resulting band intensities are converted to nucleotide insertion velocity. Kinetic fidelity studies of purified budding and fission yeast Pol δ have shown that it efficiently discriminated against misinsertion of dAMP, dGMP and dTMP opposite the template guanine. However, proofreading efficiency of mismatches was low [122,123]. Studies of Pol ε from S. cerevisiae fidelity by this method with the same sequence of template-primer demonstrated that it had similar discrimination efficiency of nucleotide misinsertion, but proofreading efficiency was higher than that of Pol δ [124].

Another method to measure fidelity is based on the ability of a Pol to fill in a single-stranded gap on M13mp2 phage DNA (Fig. 3) [125]. The single-stranded gap in M13mp2 contains the lacZ target sequence. After gap-filling by a Pol, the reaction product is introduced into competent E. coli cells and the cells are plated onto agar plates, containing the chromogenic indicator X-gal and a lawn of E. coli host cells which lack β-galactosidase activity. If the gap was replicated without any errors, the host cells acquire the ability to hydrolyze X-gal resulting in dark blue M13 plaques. Errors introduced in

Fig. 3. Outline of the gap-filling assay for DNA polymerase fidelity measurement [drawing is based on the experimental outline from 125].
gap-filling synthesis result in colorless or light blue plaques. These plaques are scored to quantify the mutation frequency. The DNA extracted from the plaques is sequenced to analyze mutations. This approach gives a broad view on replication fidelity of all 12 different substitution errors and several other types of mistake. A recent fidelity study of Pol δ, using this method, extended previous findings obtained with the gel-based method [122,123,126].

2. Aims

- To develop a system for overexpression and purification of untagged Pol ε from yeast *S. cerevisiae*.
- To characterize the structure of four-subunit Pol ε.
- To characterize the fidelity of Pol ε.
- To explore whether Pol ε and Pol δ have different biochemical properties that may help to assign their role at the eukaryotic replication fork.
3. Results

The quaternary structure of DNA polymerase epsilon from Saccharomyces cerevisiae (Paper I)

We developed an overexpression and purification protocol for untagged 4-subunit Pol ε from *S. cerevisiae* using an approach similar to earlier overexpression studies of Pol δ from *S. cerevisiae* [39]. This approach makes it possible to compare these two proteins in subsequent studies.

The *POL2* gene was subcloned into one plasmid, and *DPB2*, *DPB3*, and *DPB4* were subcloned into the second plasmid. All genes were cloned under the control of the galactose-inducible *GAL1-10* promoter. This protocol proved successful and allowed us to purify 400 µg of Pol ε from as little as 70 g of cells. The complex was purified by successive phosphocellulose, MonoQ, and MonoS chromatography and the enzyme was finally purified to homogeneity by gel filtration column or by glycerol gradient centrifugation. From a colloidal Coomassie stained gel of the peak fraction from a gel filtration column we estimated the molar ratio of subunits to be 1.00:1.12:1.07:1.06 (Pol2p:Dpb2p:Dpb3p:Dpb4p).

We determined the sedimentation coefficient for Pol ε to be 11.9 S and the Stokes radius to be 74.5 Å. The molecular mass of 371 kDa calculated from these data suggests that Pol ε is purified as a monomer of a Pol2p:Dpb2p:Dpb3p:Dpb4p complex. The frictional coefficient $f/f_0$ of Pol ε is 1.56, suggesting a moderately elongated shape of the enzyme.

To obtain a more precise molecular mass, independent of the shape of the enzyme, we carried out sedimentation equilibrium ultracentrifugation experiments. The molecular mass of Pol ε was determined to be 366 kDa under these conditions. A comparison with the predicted distribution of a monomeric versus a dimeric assembly clearly showed that Pol ε was present as a monomer in the solution under all conditions investigated.

In contrast to previously published studies where Pol ε was shown to be a dimer based on gel filtration experiments, we clearly showed that Pol ε was purified as a monomer concerning its catalytic subunit. Our value of the frictional coefficient suggested an elongated shape of the enzyme which caused an overestimation of the molecular mass determined from gel filtration experiment alone [63].

Pol ε purified as described above was used in the following papers to characterize its structural and functional properties.
Structure of DNA polymerase epsilon by cryo electron microscopy (Paper II)

Until the present study there was no structural information available on any multisubunit eukaryotic Pol. Clarification of the Pol ε structure is important to understand its biochemical properties.

We determined the structure of the 4-subunit Pol ε using cryo-electron microscopy. Images of stained and unstained Pol ε particles were aligned and analyzed using single-particle image analysis to reconstitute the three-dimensional structure of Pol ε with a resolution of ~ 20 Å. To determine the location of the largest catalytical subunit in the structure we analyzed images of purified Pol2 subunit and Pol2-Dpb2 complex preserved in the stain.

Pol ε appeared to consist of a globular domain comprising the catalytic Pol2 subunit, and an elongated tail domain, which included the Dpb2, Dpb3 and Dpb4 subunits. The two domains were joined by a flexible connection. The surface formed by Dpb2, Dpb3 and Dpb4 subunits can accommodate the double-stranded part of the primer-template DNA and position the primer end toward the polymerase and exonuclease active sites of Pol2. Using a series of primer extension assays we showed that the minimal length of double-stranded primer-template substrate should be at least 40 nucleotides. With shorter duplex DNA, Pol ε synthesized DNA rather distributively than processively, which indicates that the tail domain plays a role in maintaining a contact with the primer-template substrate. This hypothesis was supported by primer extension assays with the Pol2 subunit which synthesized DNA distributively on primer-template substrates regardless of the length of the duplex DNA.

We propose a model of the Pol ε structure with primer-template DNA, where 40 nt long duplex DNA extends from the end of the tail domain to the globular domain and the tail domain interacts with the duplex part of DNA substrate. This interaction gives a mechanical explanation to the unusually high processivity of Pol ε in absence of the clamp (PCNA).

Unique error signature of the four-subunit yeast DNA polymerase epsilon (Paper III)

To estimate the fidelity of Pol ε we purified the wild type enzyme and its mutant variant with a double amino acid changes (D289A, E291A) in the exonuclease motif of the catalytic Pol2 subunit. These amino acid changes led to a complete loss of exonuclease activity with no significant effect on the polymerase activity of the enzyme. We compared the processivity of the exonuclease-deficient Pol ε with the processivity of T4 Pol, Pol α and Pol γ on the M13mp2 DNA template. The exonuclease-deficient Pol ε showed
similar high processivity to processive T4 Pol and Pol γ, but the distribution of the product sizes appeared to be different among these polymerases, suggesting that they interact differently with the primer-template DNA substrate.

To analyze the fidelity of the wild type Pol ε we used the gap-filling assay on a M13mp2 template. The obtained mutation rates were as low as the background noise level of the assay. This places Pol ε into the group of highly accurate Pols together with eukaryotic mitochondrial Pol γ, E. coli Pol III, and Pols of bacteriophages T4 and RB69. The frequency of lacZ mutants obtained in the gap-filling assay with exonuclease-deficient Pol ε was several fold higher (6- and 13-fold in the presence of 25 or 250 μM dNTPs concentration, respectively) than that of the wild type enzyme, which is consistent with the importance of 3’ → 5’ exonuclease activity for accurate DNA synthesis by Pol ε. Sequence analysis of obtained mutants showed that the majority of errors were single-base substitutions (89% at 25 dNTP and 76% at 250 μM dNTPs), whereas single-base deletions constituted 9.3% and 18% of the mutations in both spectra, respectively. Average single-base substitution and deletion error rates were calculated to be ≥12- and ≥112-fold higher, respectively, for exonuclease-deficient Pol ε as compared to the wild type enzyme. This suggests that at least 92% of base-base mismatches and at least 99% of deletion mismatches made by Pol ε polymerase activity are corrected by its intrinsic proofreading activity.

We found that Pol ε has a specific substitution error spectrum resulting from misinsertion of dATP opposite template C, A and G and from three pyrimidine-pyrimidine mismatches, T-dCTP, T-dTTP and C-dTTP. This error spectrum was compared with the in vivo mutation spectrum in yeast strain deficient in the exonuclease activity of Pol ε (pol2-4 mutation) and in single-substitution mismatch repair (msh6 mutation). Similar to in vitro error specificity, GC → AT transitions and GC → TA transversions predominate in in vivo exonuclease-deficient Pol ε error spectrum. At the same time, the lower proportion of AT → TA transversions was determined in the in vivo spectrum than in the in vitro, and overall substitution frequency was several fold lower in vivo than in vitro.

To seek an explanation for the unique properties of the fidelity of exonuclease-deficient Pol ε we performed the amino acid sequences alignment with other B family Pols and compared those with the available crystal structure of RB69 polymerase. We found that Pol ε has non-conservative amino acids in the polymerase motif in the positions, which are conservative for all other B family Pols, and its polymerase motif contains a 66-amino acid insertion. These differences may explain the less accurate dNTP selection during DNA synthesis by the polymerase domain of Pol ε than by other B family Pols. The hypothesis that Pol ε plays a substantial role during chromosomal DNA replication is supported by observed similarities between in vivo and in vitro mutational spectra, suggesting that
Pol ε contributes to the elongation of the DNA strand at which it also proofreads replication errors. On the other hand, obvious differences in those spectra could be explained by modulation of Pol ε activity in the cell environment by additional replication factors, or by limits of the assays, used to determine in vitro and in vivo fidelities, or by limited contribution of Pol ε to replication of certain genome regions.

**Similarities and differences between the two replicative DNA polymerases, DNA polymerase δ and DNA polymerase ε (Paper IV)**

To study DNA synthesis activity and processivity of the overexpressed Pol ε we partially assembled a DNA replication system in vitro with recombinant replication factors RPA, RFC and PCNA and circular single-primed single-stranded DNA templates. In all our holoenzyme assays we compared Pol ε with Pol δ, which was purified according to an overexpression and purification protocol similar to our protocol for Pol ε [39].

We found that Pol ε was 4 times slower than Pol δ when replicating single-primed circular ssDNA. Pol δ was active and was stimulated by PCNA on both fully RPA-coated templates and on partially (20%) RPA-coated templates. DNA synthesis by Pol ε was inhibited in the presence of RPA-uncoated single-stranded template DNA and it was stimulated by PCNA only when all ssDNA was coated with RPA.

To compare the processivity of Pol ε and Pol δ we designed the holoenzyme assay to meet the criteria for single hit conditions by using a high molar ratio, 30:1, between template DNA and the Pols. Under these conditions, Pol ε was about 10 times more processive than Pol δ in the absence of PCNA and RFC on the template partially coated with RPA. Addition of PCNA and RFC stimulated processivity of Pol ε up to 6-fold and of Pol δ up to at least 100-fold. The addition of RPA to a level where all ssDNA was coated by RPA only modestly stimulated the processivity of both Pol ε and Pol δ.

From the results of the holoenzyme assays it was obvious that Pol ε and Pol δ have different properties considering the interactions with DNA and PCNA. To study the physical interactions between Pols and DNA or PCNA we used the surface plasmon resonance technique. We found that Pol ε had high affinity for primed, single-stranded and double-stranded DNA while Pol δ failed to interact with DNA. Coating the DNA immobilized on the chip with RPA decreased the affinity of Pol ε for ssDNA, but increased its affinity for primed DNA. We also showed that Pol ε failed to interact with PCNA immobilized on a chip, unlike Pol δ which interacted with PCNA with a high affinity, as demonstrated earlier [32].
Our data suggest that Pol ε and Pol δ are differently regulated while synthesizing DNA. During DNA replication in the cell RPA efficiently coats ssDNA which is produced by helicase unwinding dsDNA. This allows Pol ε to stay on the primer termini and actively synthesize new DNA strands. At the same time Pol δ synthesizes the lagging strand with the help of PCNA. However, under conditions of DNA damage, ssDNA is accumulated due to uncoupling of the helicase from the polymerase activity [127]. Phosphorylation by DNA damage checkpoint kinases causes RPA to be less efficient in the coating of ssDNA at the replication forks [128]. Pol ε would then be drawn away from the DNA synthesis and slow down the replication fork until the DNA damage was repaired and RPA is dephosphorylated.

Although Pol ε and Pol δ have comparable processivity in the presence of PCNA, the efficiency by which Pol ε synthesizes DNA is much lower then that of Pol δ. In light of described properties of Pol ε and Pol δ we suggest that the two Pols are loaded onto the primer-template junction by different mechanisms. Pol δ has a high affinity for PCNA whereas Pol ε does not have a high affinity for PCNA in solution. This may be significant for the efficiency by which Pol ε is loaded onto the primer-terminus. Slower primer utilization could explain why Pol ε is less efficient than Pol δ when replicating circular ssDNA templates.
4. Conclusions

We have developed a system for overexpression and purification of yeast four-subunit Pol ε. Recombinant Pol ε is a heterotetramer consisting of Pol2, Dpb2, Dpb3 and Dpb4 subunits at a ratio of 1:1:1:1.

We have determined the structure of Pol ε with a resolution of ~20 Å. Pol ε consists of two structurally distinct domains. A globular domain with the catalytic Pol2 subunit is flexibly connected to an elongated domain consisting of Dpb2, Dpb3 and Dpb4 subunits.

We have measured the fidelity of Pol ε in a gap-filling assay. Pol ε replicates DNA with high fidelity. The 3′-5′ exonuclease activity plays a significant role in the accuracy of DNA synthesis by Pol ε. Exonuclease deficient Pol ε generates a specific pattern of errors, not known for other DNA polymerases.

We compared the processivity of Pol ε and Pol δ in a partially reconstituted DNA replication system with a natural DNA template under single-hit conditions. In addition, we compared the DNA and PCNA affinities for both Pol ε and Pol δ. Pol ε has a high intrinsic processivity, slightly stimulated by PCNA relative to Pol δ. Pol ε and Pol δ differ from each other in their interactions with DNA and PCNA. We hypothesize that a possible function of RPA is to regulate down the DNA synthesis through Pol ε and that the mechanism by which Pol ε and Pol δ load onto a template is different due to different properties of the interaction with DNA and PCNA in solution.
Acknowledgments

I would like to thank all the people at the department of Medical Biochemistry and Biophysics for the kind and warm working atmosphere.

In addition, I would especially like to thank:

My supervisor Erik: for giving me a great opportunity to work at one of the most advanced scientific institutes. For sharing of all your invaluable experience, for all your support in my science and life. I have never wished to have a better supervisor.

Our group, Nasim, Isabelle, Göran, Else-Britt and our former members, Pawel and Helmut: thank you for being such good lab mates.

Our collaborators, especially Yury Pavlov, Polina Shcherbakova, Thomas Kunkel and Francisco Asturias: for the fruitful collaborations.

Lars Thelander and members of his group, Margareta, Vladimir and Pelle: for all your help. It was a pleasure to collaborate with you.

Ingrid, Lola and Urban: special thanks for making the work and the teaching possible.

Peter Stenlund: for teaching me the Biacore instrument.

Nick Betson: for the language check in the thesis.

Gena: for your ideas and help. You have been a good friend through the years.

Umébatucada samba school, especially Iolanda and the dancers: for the energy and good mood supply.

Our Russian community, all present and former members, Zhenya and, Misha, Valya and Kirill, Andrei and Ira, Andrei and Anna-Lena, Tanya and Andrei, Sergei, Ira, Katya: without all of you, I would feel lonely.

Моих друзей, Аню, Иру и Свету: за то, что мы вместе.

Мою семью, маму, папу и Валеру: за вашу любовь, поддержку и веру в меня.

Игоря: за твою любовь, дружбу и науку.
References


