Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze Biochimiche e Biotecnologiche

Ciclo XXVIII

Settore Concorsuale di afferenza: 05/E1 – Biochimica generale e biochimia clinica

Settore Scientifico disciplinare: BIO10 - Biochimica

Structural and kinetic characterization of DNA polymerases I and III from
Escherichia coli

Presentata da: dott. Alejandro Montón Silva

Coordinatore Dottorato	**Relatore**
Prof. S. M. Spampinato	**Prof. A. Hochkoeppler**

Esame finale anno 2015

INDEX

CHAPTER 4: DISCUSSION

Appendix I: Mass spectrometry data

Appendix II: Steady-state assays of α subunit (fittings)

Appendix III Time-traces of protein binding

Appendix IV: Time traces or real time polymerization

REFERENCES

Abbreviations

Amp	Ampicillin
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
bp	Base pairs
BS3	bis(sulfosuccinimidyl)suberate
DNA	Deoxyribonucleic acid
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycitosine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleotide
dpDNA	Double-primed Deoxyribonucleic acid
dsDNA	Double-stranded Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
HE	Holoenzyme
FRET	Förster Resonance Energy Transfer
INT	2-(4-iodophenyl)-3-(4-nitrophenyl)-5-phenyl-2H-tetrazolium
kDa	Kilodalton
KF	Klenow Fragment
mer	Affix following the number of nucleotides of an oligomer
Pol	Polymerase
PHP	Polymerase and Histidinol Phosphatase

PMSF	Phenylmethanesulfonyl fluoride
PNPase	Purine nucleoside phosphorylase
PP	Inorganic pyrophosphate
PPase	Inorganic pyrophosphatase
SDS	Sodium Dodecyl Sulfate
spDNA	Single-primed Deoxyribonucleic acid
SSB	Single-stranded binding proteins
ssDNA	Single-stranded Deoxyribonucleic acid
TRIS	Tris(hydroxymethyl)aminomethane
XOD	Xanthine oxidase

ABSTRACT

E. coli encodes 5 different DNA polymerases, although only the DNA Pol III is essential for genome replication. DNA elongation is performed by Pol III α subunit, stimulated by the association with ε and θ subunits. These three subunits define the DNA Pol III catalytic core (αεθ). There is controversy about the DNA Pol III assembly for the simultaneous control of lagging and leading strands replication, due to the fact that some Authors propose a dimeric model with two cores, whereas others have assembled *in vitro* a trimeric DNA Pol III with a third catalytic core, which increases the efficiency of DNA replication. Moreover, the function of the PHP domain, located at the N-terminus of α subunit, is still unknown. Previous studies have hypothesized a possible pyrophosphatase activity but nothing has been confirmed yet.

The present Thesis highlights by the first time the production *in vivo* of a trimeric *E. coli* DNA Pol III by co-expressing α, τ, ε and θ subunits. This trimeric complex has been enzymatically characterized and a molecular model has been proposed, defining a DNA Pol III in which 2 α subunits sustain the lagging-strand replication whereas the third core is responsible for replicating the leading strand. In addition, the pyrophosphatase activity of the PHP domain has been confirmed by a coupled-enzyme assay. This activity involves, at least, the H12 and the D19 residues, whereas the D201 should be responsible for controlling phosphate release.

On the other side, a variant of an artificial polymerase (HoLaMa), designed by deleting the exonuclease domain of Klenow Fragment, has been expressed, purified and characterized for a better understanding of bacterial polymerases mechanism. The absence of exonuclease domain impaired enzyme processivity, since this domain is involved in DNA binding. Finally, Klenow enzyme, HoLaMa, α subunit and DNA Pol III αεθ have been characterized at the single-molecule level by FRET analysis, combining ALEX and TIRF microscopy. Fluorescently-labeled DNA molecules were immobilized, and changes in FRET efficiency enabled us to study polymerase binding and DNA polymerization.

CHAPTER 1 – INTRODUCTION

I - DNA POLYMERASES

DNA polymerases establish a group of enzymes that catalyze the addition of deoxynucleotide triphosphate (dNTPs) to the 3' end of a newly-synthesized strand of DNA, using a single-stranded DNA (ssDNA) as template.

According to the Enzyme Commission (EC) classification, which was introduced during the International Congress of Biochemistry of Brussels in 1955, the polymerases are cataloged as EC 2.7.7.7, or transferases (EC 2) capable of transferring a phosphorus-containing functional group (EC 2.7), a nucleotide triphosphate (EC 2.7.7), and a deoxy-nucleotide triphosphate (EC 2.7.7.7), specifically. Figure 1.1 represents the transfer reaction catalyzed by this class of enzymes.
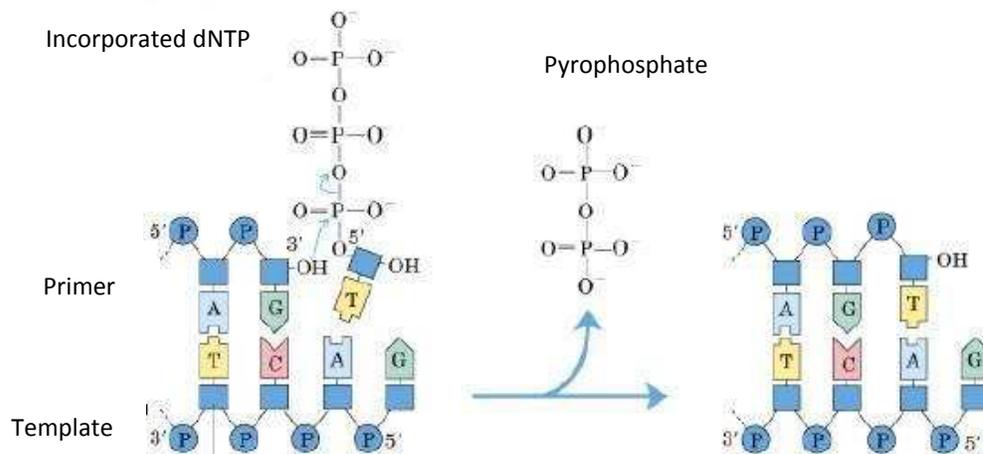
Figure 1.1: DNA polymerase reaction. A complementary dNTP is incorporated to the 3' terminus of the primer with the subsequent release of a pyrophosphate molecule. The incorporated dNTP is complementary to the template.

DNA polymerases are essential for DNA replication and repair. DNA replication is carried out by the polymerase domain, characteristic of all DNA polymerases. This domain is highly conserved at the structural level, although the amino acid sequence varies considerably among proteins. The

shape reminds an open right hand, and three distinguishable domains can be found: the Thumb domain, the Palm domain and the Fingers domain (Ollis *et al.,* 1985). Despite the fact that the catalytic site is located at the Palm domain, the other two domains are also necessary for the interaction with the substrate: DNA interacts with the Thumb domain in order to be correctly positioned towards the active site, coordinating the polymerase and exonuclease activities, whereas the Fingers domain is involved in the interaction with the incoming nucleotide and the template (Beese *et al.,* 1993a; Minnick *et al.,* 1996).

The polymerization is always in the 5' → 3' direction, promoting the nucleophilic attack to the α-phosphate of the nucleoside triphosphate to be added. Currently, there are no data about the existence of any polymerase able to catalyze the reaction without the aid of a DNA or a RNA primer. Hence, depending on the primer used, they are defined as DNA-dependent or RNA-dependent polymerases.
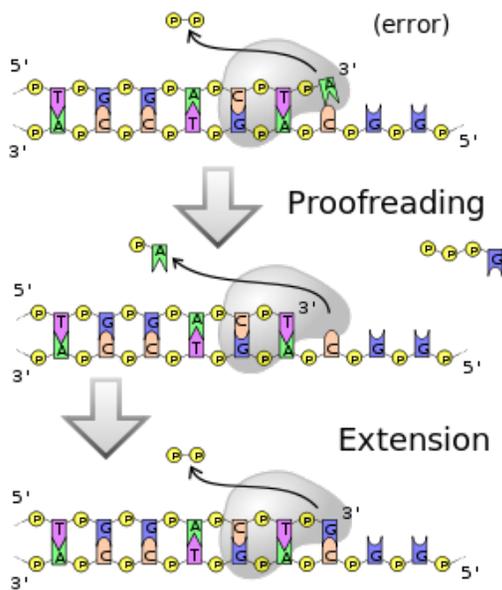
Figure 1.2: Scheme of proofreading activity.

The correction of errors involved in replication is, instead, in charge of a second domain with 3' → 5' exonuclease activity, present in some, but not all, DNA polymerases. Generally, in Gram positive bacteria, this activity resides in a polymerase-adjacent domain, whereas in Gram negative bacteria the exonuclease activity takes part in independent subunits of the polymerase (Kelman & O'Donell, 1995). In any case, if a non-complementary nucleotide is integrated during polymerization, the exonuclease active site removes the wrong nucleotide hydrolyzing the phosphodiester bond. Once removed, the polymerase domain adds a new nucleotide at that corrected site and goes on with the polymerization process. Some DNA polymerases, as DNA Pol I, possess a 5' → 3' exonuclease domain as well, essential for removing primers generated during the replication of the genome.

1.1 Functional properties of DNA polymerases

In spite of being quite diverse, DNA polymerases share a group of features present in almost all these enzymes (Kornberg & Baker, 1992):

- Directional synthesis of DNA, since DNA polymerases catalyze the synthesis in the 5' → 3' direction.

- Use of dNTPs to carry out the DNA polymerization process by sequential incorporation of a complementary nucleotide, with the subsequent release of a molecule of inorganic pyrophosphate (PP).

- Necessity of a template: The polymerization reaction is developed by a DNA template that is copied according to the base pairing rules predicted by Watson and Crick (1953), thereby establishing the order of incorporation of dNTPs. This finding was particularly important, not only for providing the chemical semiconservative basis of DNA replication, but also because it represented the first example of template used for guiding a biosynthesis reaction.

- Necessity of a primer: To initiate the DNA synthesis, a primer which brings an -OH group is necessary for nucleotides incorporation, since they cannot perform the *de novo* synthesis of a DNA chain. An initiator can be a segment of DNA or RNA chain.

- Presence of a pair of divalent ions, Mg^{2+} preferably (Steitz *et al.,* 1994), to act as cofactor of the reaction process and allow the transfer of nucleotides.

Obviously, not all polymerases are identical and some functional differences can be found between them, as the presence of typical activities such as 5' → 3' exonuclease activity in DNA Pol I (Kornberg & Baker, 1992) or SOS response of DNA polymerases IV and V (Woodgate & Ennis, 1991).

II - DNA POLYMERASES IN *E. coli*

Table 1 summarizes the five DNA polymerases found in *Escherichia coli.*

Enzyme	Family	Description
DNA Pol I	A Family	One polymerase domain and two exonuclease domains
DNA Pol II	B Family	DNA repair
DNA Pol III	C Family	Complex holoenzyme. The essential polymerase in *E. coli*
DNA Pol IV	Y Family	Involved in SOS response and TLS
DNA Pol V	Y Family	Involved in SOS response and TLS

Table 1: Classification of DNA polymerases in *E. coli*

Among the five polymerases present in *E. coli*, only the DNA Pol III is considered to be essential. DNA Pol III consists of 10 different types of subunits and it is capable of replicating the entire chromosome of the bacterium in 40 minutes. DNA Pol III was found out in 1971 (Kornberg & Gefter, 1971), almost 20 years later than the DNA Pol I. (Bessman, 1956). The use of an *E. coli* mutant strain missing the DNA Pol I (De Lucia & Cairns, 1969) was essential for the identification of the DNA polymerases II and III. Due to the absence of DNA Pol I, it was possible to characterize these polymerases that, compared with DNA Pol I, are considerably less expressed. The DNA polymerase II and the DNA polymerase III are present at the level of 50 and 20 copies per cell (McHenry & Kornberg, 1997), respectively, while 400 copies per cell were estimated for the DNA polymerase I (Kornberg & Baker, 1992). Nevertheless, DNA Pol III appears to have an extraordinary efficiency and speed of polymerization, integrating approximately 1000 nucleotides per second (Maki *et al.,* 1988).

On the other hand, DNA Pol II is required for the repair of the DNA gaps and is able to polymerize up to 100 consecutive nucleotides in the 5' → 3' direction. It features high fidelity of replication, but at the same time features low processivity, which makes it unsuitable for genome replication (Banach-Orlowska *et al.,* 2005). Pol IV and V are the most recently discovered polymerases in *E. coli* (Wagner *et al.,* 1999; Tang *et al.,* 1999). Both polymerases are expressed as SOS response to DNA damage and they are involved in the translesion-synthesis (TLS) (Jarosz *et al.,* 2007). All polymerases are described in detail below, especially DNA polymerases I and III, subject of this Thesis.

1.2 DNA polymerase I

The DNA polymerase I (103 kDa), encoded by the *polA* gene, is the most abundant polymerase in *E. coli*, despite not being primarily responsible for the replication of the parental DNA. It is a single polypeptide chain that, by proteolysis, can be split into two main fragments, each one with different particular functions: the largest one (Klenow fragment, 68kDa) features 5' → 3' polymerase activity and 3' → 5' exonuclease activity; the other one, smaller (35kDa), is outfitted with 5' → 3' exonuclease activity (Klenow & Hennigsen, 1970).

The DNA Pol I of *E. coli* is considered to be the first discovered polymerase, (Bessman, 1956), with about 400 copies per cell (Kornberg & Baker, 1992). This enzyme performs both replication and DNA repair functions. During replication, DNA Pol I, together with RNase H, is fundamental so as to remove the RNA primers on the lagging strand after the synthesis of Okazaki fragments, and to replace them with suitable DNA sequences, mimicking the role of a reverse transcriptase (Riccheti & Buc, 1993); with regards to the proofreading function, DNA Pol I is able to recognize mismatches or damaged DNA (i.e. UV radiation damage) and repair it (Tait *et al.,* 1974). In addition, DNA Pol I is the main polymerase involved in the ribonucleotide excision repair (RER) system in *E. coli* (Vaisman *et al.,* 2014).

It is noteworthy that the homology of the amino acid sequence of the DNA pol I in different microorganisms (i.e. *E. coli, Thermus acquaticus* and *Chlamydia trachomatis*) is approximately 35% (Patel *et al.,* 2001). This homology is principally located in six recurring structural motifs involved in DNA binding, dNTPs binding and conformational movements. In spite of this low sequence identity, the structures of DNA Pol I of different organisms are highly superimposable, suggesting an identical post-translational folding (Fig. 1.3).

The *in vivo* confirmation of this finding was provided by expressing the Taq Pol I of *Thermus acquaticus* in *E. coli* strains, lacking the endogenous DNA Pol I. This deletion normally triggers lethal effects, which, however, was solved by the expression of Taq Pol I (Suzuki, 1996). Several experiments of site-specific mutations and/or random mutagenesis in the highly conserved sequences have revealed that there are three irreplaceable residues (Asp610 for binding to Mg^{2+}, Arg659 and Lys663 for the binding to dNTPs); in contrast, other residues tolerate conservative substitutions (Tyr611 and Glu615 to stabilize the tertiary structure; Phe667, Gly668 and Tyr671 for

binding to dNTPs), while the remaining residues are likely to be replaceable without affecting the enzymatic activity. However, other two important parameters were altered: the specificity towards the substrate and the fidelity of replication (Patel & Loeb, 2000).
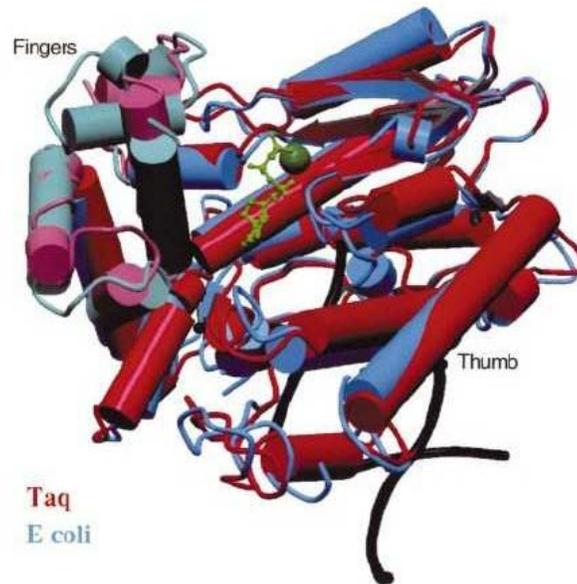
Figure 1.3: Superposition of crystal structures of Taq Pol I (in red) and *E. coli* DNA Pol I (in blue). Only the region corresponding to the Fingers domain (in magenta) differs significantly between both enzymes.

As it was previously mentioned, the DNA Pol I of *E. coli* is encoded by the *pol*A gene, and is formed by 928 amino acids, featuring a theoretical molecular mass equal to 103 kDa (De Lucia & Cairns, 1969). Its structure is divided in 3 different functional domains:

- 5' → 3' exonuclease domain (residues 1-323)
- 3' → 5' exonuclease domain (residues 324-517)
- 5' → 3' polymerase domain (residues 521-928)

The latest two domains, above-mentioned, constitute the well-known "Klenow fragment", described more in detail below, whose biotechnological applications are being developed nowadays.

1.2.1 Klenow fragment

In 1970, Klenow and Henningsen had noticed that during the purification of DNA polymerase I (then, the only known polymerase *in E. coli*) there were two different peaks of polymerase activity: one of them corresponded to a 70 kDa protein with low exonuclease activity (enzyme A), and another one corresponded to an enzyme of 109 kDa with normal exonuclease activity (enzyme B). Treating the population B with subtilisin protease, authors obtained a single peak of polymerase activity in gel filtration. The isolated enzyme featured a molecular mass equal to that of the population A. Both researchers concluded that the population A derived from a limited proteolysis of population B, discovering what is currently known as Klenow fragment (Klenow & Henningsen, 1970). Numerous structural studies of the Klenow enzyme have been conducted, (unlike completed DNA pol I that has not been fully crystallized yet), and it was therefore possible to study the active site of the Fragment as well as the binding sites with substrates (Freemont *et al.,*1988; Beese *et al.,*1993b; Li *et al.,*1998).

Figure 1.4: Klenow Fragment from *E. coli* (PDB 1KFD) complexed with dNTP and pyrophosphate molecule (Beese *et al.,* 1993a).

The Klenow Fragment and the DNA polymerase I are two of the most popular enzymes used in biotechnology, due to being among the first discovered polymerases. The removal of the 5' → 3' exonuclease domain made the DNA Pol I suitable for many scientific applications. Consequently, the Klenow Fragment is currently used in:

- Synthesis of double-stranded DNA from a single strand.
- DNA labeling with radioactive nucleotides (Nick translation): in the presence of a nick in a double-stranded DNA, the DNA pol I removes some of the nucleotides (Hours & Denhardt, 1979) and replace them with their labeled analogues (Fig. 1.5), creating a labeled DNA sequence that can be used in Fluorescence in situ Hybridization (FISH) as well as in radiolabeling.
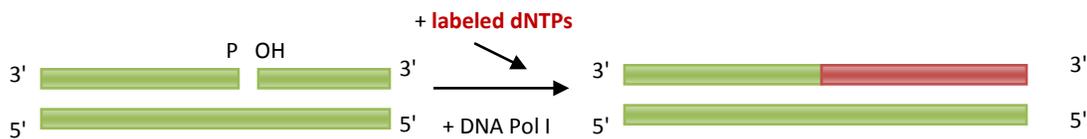
Figure 1.5: Nick translation and generation of a labeled DNA sequence (shown in red) by the different domains of DNA Pol I.

- Production of non-cohesive ends (blunt ends) from cohesive ends (sticky ends), using the polymerase domain to extend the strand in the 5' → 3' sense (Yang *et al.,* 2005) (Fig. 1.6.A), or by using the exonuclease domain to shorten the protruding filament in the 3' → 5' direction (Zhao *et al.,* 2013) (Fig. 1.6.B).
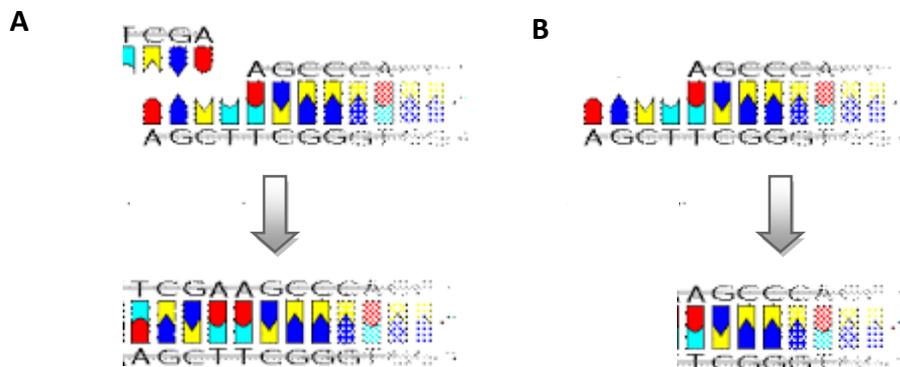
Figure 1.6: Production blunt ends from sticky ends by the action of the polymerase activity **(A)** or by the proofreading activity **(B)** of Klenow enzyme.

In addition, the Klenow fragment was initially used by Kary Mullis in the Polymerase Chain Reaction (PCR), before being replaced by the thermostable Taq polymerase (Saiki *et al.,* 1986). Furthermore, Klenow fragment has been subject of study at the single-molecule level. For instance, by using a single-molecule FRET approach, it was shown that mismatched dsDNA binds to the polymerase in a different disposition than in the DNA with a correct nucleotide complementarity (Markiewicz *et al.,* 2012). Furthermore, in the presence of a correct dNTP, the nucleotide stabilizes the DNA at the *pol*-site, making the DNA-polymerase-dNTP complex progress to the closed disposition. This conformational change with the correct dNTP is so fast that the intermediate (*ajar)* conformation is scarcely appreciable. However, in the presence of a mispaired primer terminus, the polymerase should invert to the *ajar* conformation, with the subsequent DNA switching to the *exo* site so as to exert the proofreading activity (Berezhna *et al.,* 2012). Thus, Klenow Fragment can discriminate between correct and incorrect nucleotides, being essential to the fidelity of DNA replication.

1.3 DNA polymerase II

The DNA polymerase II (89.9 kDa) (Fig. 1.7) is encoded by the *polB* gene, being a high-fidelity replicative enzyme with 3' → 5' exonuclease activity in its N-terminal portion. The number of copies in *E. coli* is 30-50 per cell. DNA Pol II was discovered in 1970 (Knippers, 1970) and it is considered the prototype for the B Family DNA polymerases (Bonner *et al.,* 1990), since it contains the sequence of a highly conserved motif in different DNA polymerases of different kingdoms.
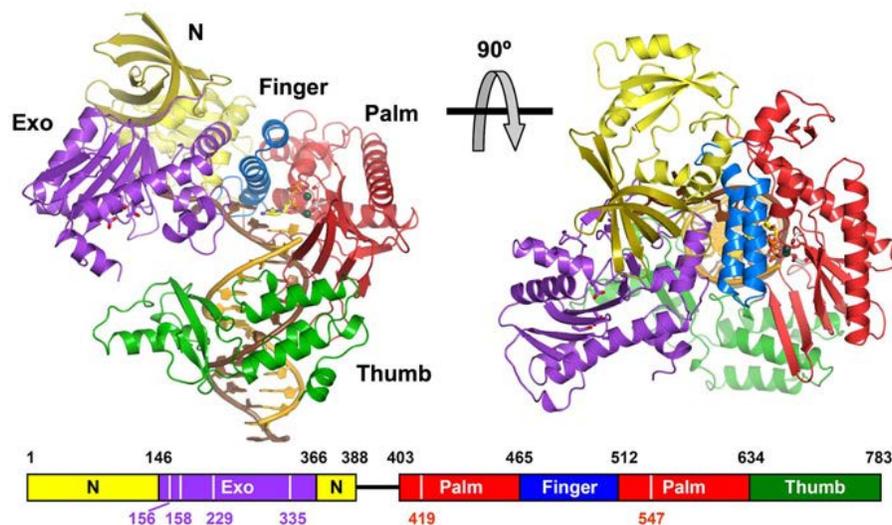
Figure 1.7: Crystal structure of *E. coli* DNA Pol II complexed with DNA and dCTP (Wang & Yang, 2009).

As a result of its 3' → 5' exonuclease activity, DNA Pol II carries out a faithful replication mechanism: the enzyme is able to synthesize DNA with an error rate less than 10^{-6}. In contrast, a mutant variant enzyme, deficient for this exonuclease activity, showed a frequency of replication errors 13-240 times higher than the wild type DNA Pol II (Cai *et al.,* 1995).

DNA Pol II plays an important role in *gap filling*, completing the single-stranded regions that block the intervention of the DNA polymerase III, without needing a RNA primer. Moreover, the activity of Pol II is stimulated by the presence of single-stranded DNA binding proteins (SSB), responsible for binding to ssDNA and protect it against spontaneous annealing of individual strands.

DNA Pol II significantly helps DNA Pol III during DNA replication (Banach-Orlowska *et al.,* 2005; Fijalkowska *et al*., 2012), due to the fact that the DNA Pol II contacts the β-clamp and the loading clamp of the DNA Pol III (Hughes *et al.,* 1991; Heltzel *et al.,* 2009a). Even though an important number of mismatches will be removed by the ε subunit of DNA Pol III, some of the mismatches could remain since there are some points in which DNA Pol III may stall, such as at particular DNA lesions. In that case, there is a critical point for the polymerase because the possibility of mismatches conversion into potential mutations increases. Fortunately, the proofreading activity of Pol II solves this problem by removing replication errors and thus acting as an extrinsic proofreader of the DNA Pol III. In parallel, the possibility that another polymerase may have access to the replication fork, correcting errors from other polymerases, has already been demonstrated in other microorganisms. That is the case of *Saccharomyces cerevisiae* (Pavlov *et al*., 2004), in which a mutant variant of DNA Pol ε (Y831A) presented a lower proofreading activity but without severe consequences in cell viability, supposing that other polymerases contribute to error removing.

Finally, several studies indicated that the DNA Pol II is involved in other DNA repair procedures, including DNA repair from stress-induced mutagenesis (Hastings *et al.,* 2010), intrastrand cross-links (Berardini *et al.,* 1999), DNA damaged by oxidation (Escarceller *et al.,* 1994) or UV irradiation (Rangarajan *et al.,* 1999).

1.4 DNA polymerase III

The DNA polymerase III of *E. coli* is the main replicative system, responsible for the faithful replication of the entire genome of the bacterium, which is around of 4.6 Mbp (Maki *et al.,* 1988). The holoenzyme polymerase III (Pol III HE) is in fact defined as the association between the cores (formed by α, ε and θ subunits) a clamp loader (formed by the assembly of τ, γ, δ, δ', χ and ψ proteins) and the β2 homodimer-sliding clamp (McHenry, 2011). Besides, the holoenzyme associates with other proteins with ancillary functions for replication, such as the DnaB helicase, which unwinds the double helix of DNA in order to provide a single chain template to the polymerase; the DnaG primase, which synthesizes a short 12 nucleotides-RNA primer to provide the polymerase the 3'-OH end, so as to initiate the replication process; and the SSB proteins, which prevent degradation of single-stranded DNA.

The DNA polymerase III of *E. coli* was discovered in 1971 (Kornberg & Gefter, 1971). DNA polymerase III is present at the level of 20 copies per cell (McHenry & Kornberg, 1997), considerably lower than the 400 copies per cell of DNA Pol I. Table 2 summarizes the distribution of the subunits that define the Pol III HE.

Structural complex	Subunit	Encoding gene	Molecular mass (kDa)
Catalytic core	α	*dnaE*	129.9
	ε	*dnaQ*	27.5
	θ	*holE*	8.6
Clamp loader	τ	*dnaX*	71.1
	γ	*dnaX*	47.5
	δ	*holA*	38.7
	δ'	*holB*	36.9
	χ	*holC*	16.6
	ψ	*holD*	15.2
Sliding clamp	β	*dnaN*	40.6

Table 2. Structural distribution of different subunits of *E. coli* DNA Pol III HE

In spite of being the polymerase subunit, α subunit is not able by itself to achieve the huge degree of fidelity of replication observed *in vivo*. Therefore, only through the association with other factors of the holoenzyme complex, α subunit reaches the highest speed (1000 bp per second) (Maki & Kornberg, 1988) and fidelity of replication. In addition, there are still some important aspects about DNA Pol III HE that have not been determined completely, as the crystallization of some of its subunits, the role of some of its sub-domains or the characterization at the single-molecule level, due to the high complexity of the holoenzyme.

1.4.1 Catalytic core ($\alpha\epsilon\theta$)

The catalytic core of DNA Pol III HE is a heterotrimer system composed of α, ϵ and θ subunits, encoded *by dnaE* (Welch *et al.,* 1982) *, dnaQ* (Horiuchi *et al.,* 1981) and *holE* genes (Studwell-Vaughan & O'Donnell, 1993), respectively. This core is a very stable complex with a linear α-ϵ-θ order and a 1:1:1 ratio of the three mentioned subunits (Studwell-Vaughan & O'Donnell, 1993). Additionally, there are interactions between α, ϵ and θ subunits: the α subunit interacts through its N-terminus with ϵ (Maki & Kornberg, 1987), which in turn contacts θ through its N-terminal region (Conte *et al.,* 2012).

The α subunit contains the polymerase active site (Maki *et al.,* 1985), whereas the ϵ subunit is responsible for the 3' $\rightarrow$ 5' proofreading activity (Scheuermann *et al.*, 1984). In contrast, enzymatic activity has not been identified yet for the θ subunit (Studwell-Vaughan & O'Donnell, 1993), but it is important for the stabilization of the general form of the core. The three core subunits cooperate and encourage each other. For example, the association of ϵ and α subunits increases the exonuclease activity from 10 to 80 times (Maki & Kornberg, 1987), whereas the interaction between ϵ and θ increases both the stability and the exonuclease activity of ϵ (Studwell-Vaughan & O'Donnell, 1993), being ϵ a critical factor in the assembly of the core. This role would be strongly influenced by the θ subunit, preventing ϵ subunit degradation (Conte *et al.,* 2012). At the same time, ϵ appears to have a 3-fold stimulatory effect on α subunit processivity (Maki & Kornberg, 1987).

1.4.1.1 α subunit

The α subunit (129.9 kDa), encoded by *dnaE* gene and composed of 1160 amino acids, features the 5' → 3' polymerase activity, and is shaped as an "opened-right hand", typical of all DNA polymerases. α subunit shows a clear differentiation of three structural domains: Palm (residues 271-432 and 511-560), where the polymerase active site is found; Thumb (residues 433-510), involved in DNA movements and binding during catalysis; and Fingers, which interacts with the incoming nucleotides (Brautigam & Steitz, 1998). The latest can be divided in Index (residues 641-756), Middle (541-640 and 757-778), Ring (779-838) and Little Finger (839-911) sub-domains. Moreover, there is a PHP domain (first 270 residues of α subunit), below explained, located at the N-terminal portion of the protein.
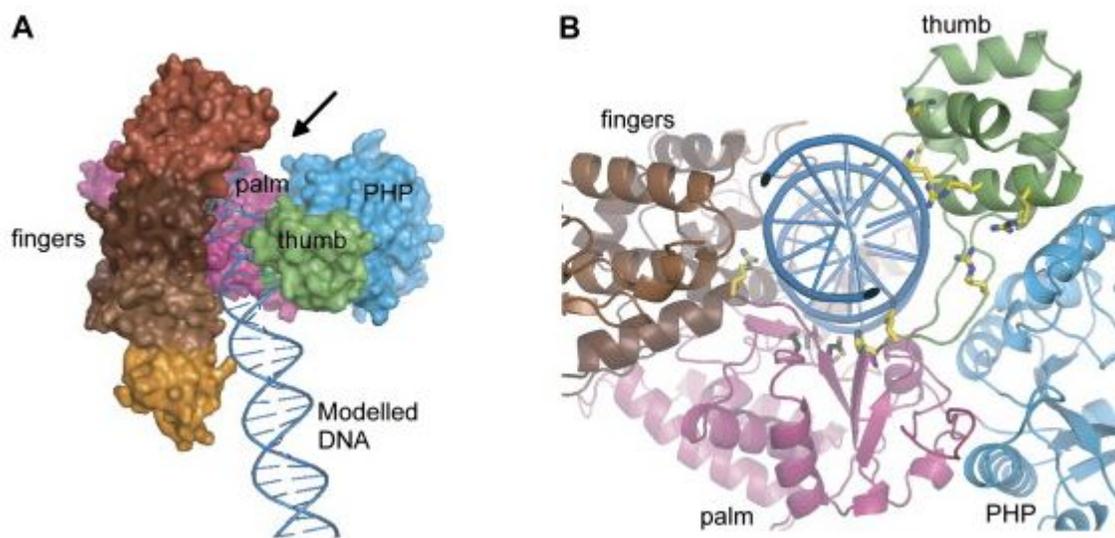
Figure 1.8: (A) Surface model of *E. coli* DNA Pol III α subunit complexed with DNA(Lamers *et al.,* 2006). **(B)** Detail of the groove in which the DNA is located; positively charged residues, which interact with the DNA backbone, are highlighted in yellow (Lamers *et al.,* 2006).

Anyway, the structure of α subunit has never been determined yet, due to its size and complexity. It was just partially resolved by X-ray diffraction (Lamers *et al.,* 2006) (Fig. 1.8). Indeed a fragment comprising residues from 1 to 917 (1160 residues in total), being absent the C-terminal portion where the area of binding to τ subunit is located. Palm, Thumb, and Fingers domains create a deep cleft that allows dsDNA to fit and interact with the active site, located at the Palm domain.

The PHP domain is found close to the Thumb domain, whereas a GS motif, also conserved in β polymerase, is present in α subunit and is formed by glycine 363 and serine 364, creating a hydrogen bond with the tail of phosphates of the incoming nucleotide, which might bind to a loop that connects the Index and the Middle sub-domains, at the domain of the Fingers.

McCauley *et al.,* (2008) determined the presence of a Helix-hairpin-Helix (HhH) sequence, repeated twice in tandem in the 833-889 residues region (Fingers domain) and included in the area of binding to the β subunit. In addition, an oligonucleotide binding (OB-fold) domain in the region of residues 964-1078 was also identified. These last two domains are able to bind the dsDNA in a non-specific way, and the ssDNA at the strictly conserved F1031 residue, respectively.
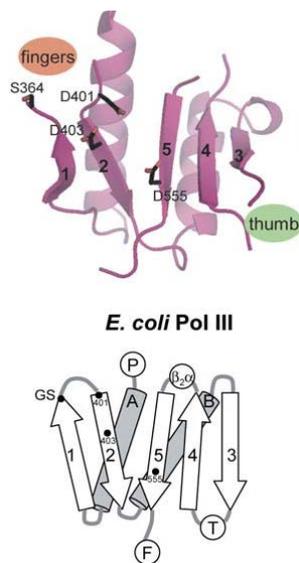
Figure 1.9: Structure of the polymerase active site of *E. coli* DNA Pol III α subunit (Lamers *et al.,* 2006).

The active site consists of a five filaments β sheet, arranged in the 1-2-5-4-3 order, where the symmetry is interrupted by the antiparallel central filament (Fig. 1.9), as in the case of the polymerase β of the X family polymerases (Davies *et al.,* 1994). Moreover, the active site of α subunit features three conserved aspartate catalytic residues, known as catalytic triad, (D401, D403 and D555) (Pritchard & McHenry, 1999). While the first two catalytic aspartates are placed at the second filament, the third one is located at the fifth filament at a distance of 152 residues.

The mechanism of polymerization involves the nucleophilic attack to the α-phosphate of an incoming dNTP, according to the base pairing rules of Watson and Crick (1953). The process goes ahead with the formation of a phosphodiester bond and the subsequent release of a molecule of inorganic pyrophosphate (PP), being a common reaction to all known polymerases (Joyce & Steitz, 1994).

The model proposed by Brautingham *et al.,* (1998) established the presence of two divalent Mg^{2+} ions (A and B), coordinated by aspartate residues in the active site. On the one hand, the metal ion A is coordinates the nucleophilic attack to the α-phosphate of the incoming nucleotide, raising the acid dissociation constant of the hydroxyl group of the primer to facilitate its deprotonation during the nucleophilic attack. On the other hand, the metal ion B is placed between the incoming dNTP and one catalytic aspartate, stabilizing the outgoing pyrophosphate, produced by the polymerase reaction, and orientating the phosphate tail of incoming nucleotides for the catalysis. Moreover, metal B controls the negative charge, developed during the pentavalent transition state formation. In sum, this model considers that the metal A prepares the nucleophilic attack of the incoming nucleotide α-phosphate, whereas metal B stabilizes the outgoing pyrophosphate.

Based on the above-mentioned partial crystallization data of α subunit (Lamers *et al.,* 2006) and the analysis of the structure of the analogous protein from *T. aquaticus* (Wing *et al.,* 2008), it was possible the study of the structure of *E. coli* α subunit and its modifications in the presence of DNA, assuming a folding around the double helix of DNA, positioned into the cavity located between the three main domains.

Initially, ssDNA template comes into α polymerase controlled by the Fingers domain (point 1, fig. 1.10), directed toward the Palm domain (point 2, fig. 1.10). The Thumb domain (point 3, fig. 1.10) closes around DNA, moves about 5° and contacts with the DNA at the minor groove by two positively charged α helices (residues 437-461). While the Fingers domain rotates 15° approximately towards the DNA, allowing the formation of an input channel for the incoming dNTP, the C-terminal region rotates about 20°, placing the binding site for the sliding clamp close to the newly synthesized dsDNA. At the end of nucleotide incorporation, the neo-synthesized dsDNA contacts the Thumb domain (point 3, fig. 1.10) along its minor groove while the polymerase flows on the template. Subsequently, the DNA interacts with the extended end portion of the Fingers (point 4, fig. 1.10) and enters the ring formed by the β clamp, located at the point 5 in fig. 1.10.
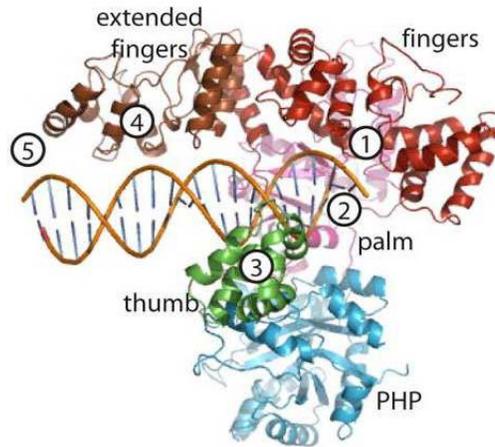
Figure 1.10: Model of α subunit bound to DNA (Lamers & O'Donnell, 2008); numbers explained in the text.

Additionally, a conserved loop (residues 753–758 in Pol III) between the index and middle Finger subdomains comes close to the 3'-terminal nucleotide. This loop contains two aromatic amino acids (Tyr754 and Phe756) that interact with the emergent base pair. Besides, a cluster of arginine, exposed at the Palm (R390 and R396) and at the Fingers (R709 and R710) domains, interacts with the γ phosphate of the incoming nucleotide (Lamers *et al.,* 2006) (Fig. 1.11). In the Palm domain, another highly conserved residue (lysine 553) is also present and is thought to be involved in substrate positioning. K553 stabilizes the phosphate at the 5' end of the last base of the extended filament (McHenry, 2011).
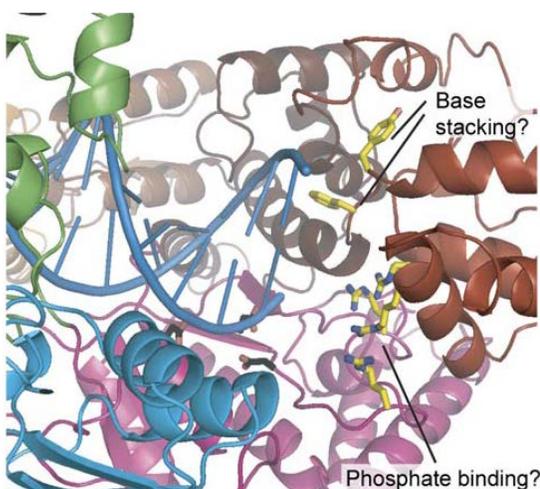
Figure 1.11: Polymerase active site of α subunit with the catalytic triad (D401, D403 and D555) and the residues presumably involved in dNTP contact (R390, R396, R709 and R710) highlighted in black and yellow, respectively (Lamers *et al.,* 2006).

16

The proposed model hypothesized that DNA contacts the active site of α subunit with an angle of 30°, approximately, respect to the orientation of the β sheet. This inclination differs from the bond observed in the crystal structure of *Thermus aquaticus* DNA Pol III α subunit (Wing *et al.,* 2008), where DNA featured a parallel disposition respect to the β sheet orientation. This difference could be fundamental and could indicate the existence of different ways of binding the substrate among DNA polymerases from family C. In contrast, the proposed conformation of the bond and the observed angle is very similar to the observations in the DNA Pol C of *Geobacillus kaustophilus* (Evans *et al.,* 2008). The similarities between the structures of DNA Pol C complexed with the substrate, and that of *E. coli* DNA Pol III α subunit devoid of DNA (Fig. 1.12), makes possible to assume that the domains of the Palm and Fingers should undergo large conformational changes closely dependent on the substrate binding, as occurs in the A and B polymerase families (Steitz, 2006). In the case of DNA polymerases from family C, movements might be more extensive and accompanied by a change of the Palm domain disposition, which may rearrange the catalytic residues modifying the catalytic site.
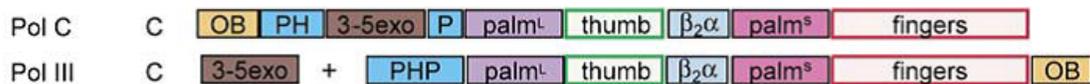
Figure 1.12: Domain organization of DNA Pol C and *E. coli* DNA Pol III (Lamers *et al.,* 2006).

Since the DNA polymerase III is responsible for genome replication in *E. coli*, mutations in its *dnaE* gene produce strains featuring a mutator phenotype, especially when mutations are located in the Palm, Fingers and Thumb domains (involved in the catalytic reaction, discrimination of the incoming nucleotides and substrate binding, respectively), instead of in the PHP domain or in the area of binding between α and β subunits (Vandewiele *et al.,* 2002).

1.4.1.1.1 PHP domain

i) Generalities and evolution in microorganisms

The PHP domain is considered to be one of the most conserved domains among different bacterial polymerases (Timinskas *et al.,* 2014). The PHP (Polymerase and Histidinol Phosphatase) superfamily was identified in 1998 through sequence alignments, based on the conservation of four motives containing histidines and aspartates in suitable positions for the coordination of metal ions (Aravind & Koonin, 1998) (Fig. 1.13).

The conservation of the PHP domain suggests the importance of the role which plays in DNA replication. Moreover, the highlighted sequence analogy between the PHP domain and other phosphoesterase proteins permitted to suggest its involvement in hydrolysis or in the release mechanism of the pyrophosphate, produced by the polymerase reaction (Aravind & Koonin, 1998). This kind of activity, not yet described, would promote a constant displacement of the polymerase reaction balance, thus being crucial in the processivity of the DNA Pol III.

The structural conservation level is significantly high, both in the PHP domain of the C polymerases of Gram positive bacteria like *G. kaustophilus* (Evans *et al.,* 2008) and in replicative polymerases of different Gram negative bacteria. In particular, the PHP domain consists of four conserved regions (motifs I–IV) (Aravind & Koonin, 1998) (Fig. 1.13), identified by nine residues (histidines and aspartates), which form a coordination site for metal ions and are likely to be involved in catalysis. Motif I features a dyad of histidines, separated by a single residue; motifs II-IV contain additional conserved histidines and aspartates, which may thus be involved in catalysis by metal coordination. Motif IV, however, is missing in some *archaeal* and bacterial proteins, suggesting the possibility that PHP family proteins may differ in the number of coordinated metal ions.
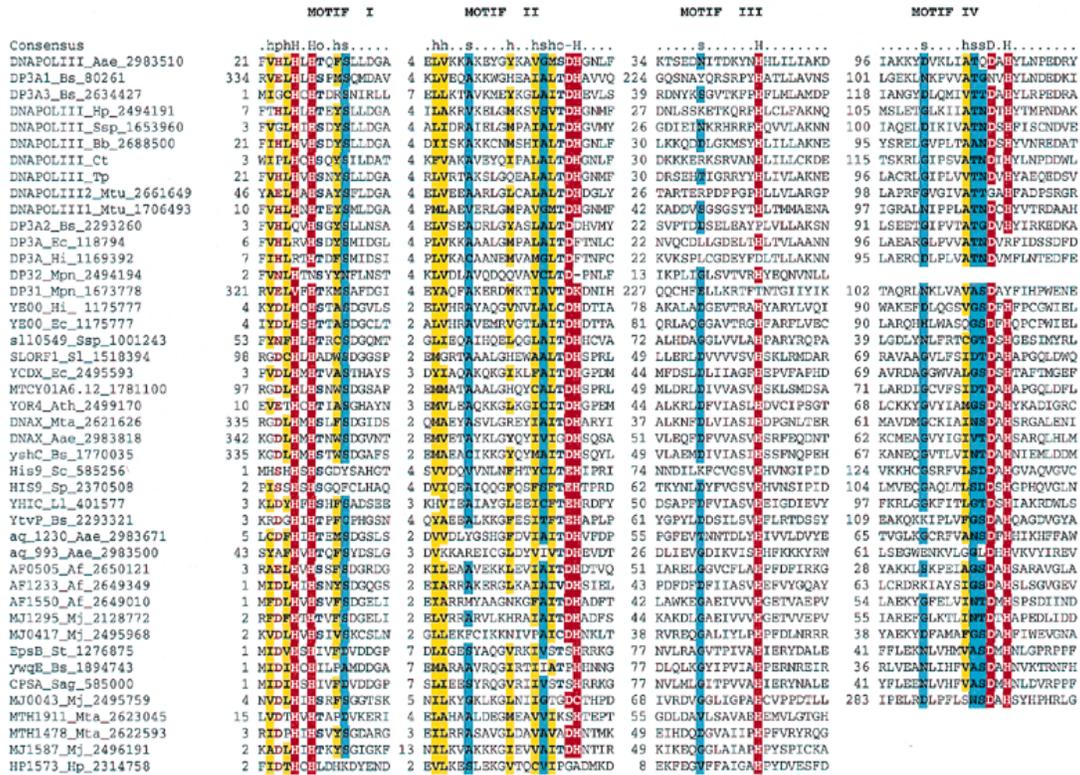
Figure 1.13: Sequence alignment of proteins from the PHP superfamily (Aravind & Koonin, 1998).

The PHP domain in DnaE protein of *T. thermophilus* (Stano *et al.,* 2006) and *T. aquaticus* (Wing *et al.,* 2008) and the PHP domain of Gram positive *B. subtilis* polymerase X (Baños *et al.,* 2008) are characterized by the extreme conservation of all nine amino acids of the coordination cluster. In some polymerases, the PHP domain has been associated with a 3' → 5' exonuclease activity, divalent metal ions-dependent, as occurs in the case of the Mn^{2+} dependent-DNA polymerase X of *B. subtilis* (Baños *et al.,* 2008) and in the case of the α polymerase of *Thermus thermophilus*, which requires Zn^{2+} and catalyzes the same exonuclease activity than ε subunit (Stano *et al.,* 2006). In the same domain of *B. subtilis,* an intrinsic apurinic/apyridinic 3' → 5' endonuclease activity, Mn^{2+}-dependent, has also been identified (Baños *et al.,* 2010). In contrast, DnaE protein of *E. coli* does not present the mentioned activity, since it is performed by the proofreading ε subunit. In other cases, deletion experiments, as those conducted on the DNA polymerase X of *D. radiodurans,* have shown that the PHP domain plays an important structural role for maintaining the 3' → 5' exonuclease activity and the proper folding of the protein (Blasius *et al.,* 2006).

19

These differences, related to the conservation of the exonuclease activity, internal or external to the replicase, allowed to hypothesize the evolution of the various forms of replicases from a common ancestor (Huang *et al.,* 1997). The evolution of this ancestor would have led to systems where the polymerase is associated with independent exonuclease protein (as α and ε subunits in DNA Pol III of *E. coli*), or individual proteins bearing both polymerase and exonuclease activities (like polymerase III of thermophilic organisms and polymerase C of Gram positive bacteria).

DnaE protein, depending on the particular microorganism, may employ two different types of exonuclease activities: one linked to the established interaction with the ε subunit, and another one associated with the PHP domain activity. Actually, DnaE proteins with a PHP containing all the residues of the coordination cluster have been identified as precursors of proteins that possess a PHP missing these residues. Therefore, the model highlighted the possibility that the exonuclease activity, originally in charge of the PHP domain, has been supplanted because of the interaction between α and ε subunits (Barros *et al.,* 2013).

The structure of the PHP domain was resolved for proteins belonging to different species, as the Histidinol phosphate phosphatase HB8 of *Thermus thermophilus* (Omi *et al.,* 2007) and the YcdX protein of *Escherichia coli* (Teplyakov *et al.,* 2003), and it was configured as a distorted $\alpha_7\beta_7$ TIM barrel. In databases, structural information about the family X of DNA polymerases is also available, as in the case of *D. radiodurans* Pol X (Leulliot *et al.,* 2009), in which the PHP domain is located at the C-terminus of the protein. There is also information from the C family polymerases, where the PHP domain is located at the N-terminal region, like in α polymerase of *Thermus aquaticus* (Bailey *et al.,* 2006) and *E. coli* (Lamers *et al.,* 2006). A typical feature of the polymerases from family C is the orientation assumed by the β4 strand, which is antiparallel to the rest of the strands. The structure of the PHP domain of the PolC polymerase of *Geobacillus kaustophilus* contains a 3' → 5' exonuclease activity domain inserted in the loop that connects β3 and β4 filaments (Evans *et al.,* 2008) (Fig. 1.14).
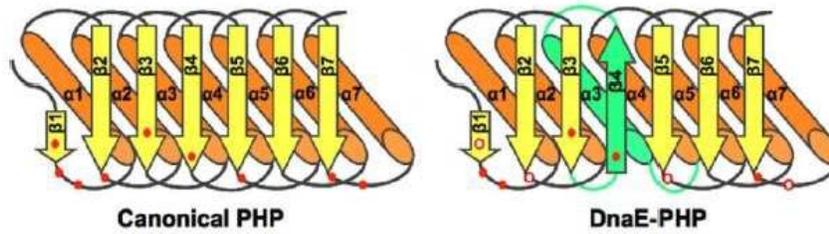
Figure 1.14: Structure of the canonical PHP and the DnaE-PHP domain barrels. Filled red circles indicate conserved metal-chelating residues in all PHP domains; open circles indicate metal-chelating residues in PolC and Taq DnaE but not conserved in *E. coli* DnaE (Evans, *et al.,* 2008).

The aspartate and histidine residues, capable of coordinating metal ions, were identified in the conserved motifs, in correspondence of the C-terminus of the β strand, which is the expected canonical position of the active site in the αβ barrels. The coordinated metals within the active site are usually three, but in *Proteobacteria* polymerases, such as *E. coli, Rickettsia* or *Haemophilus influenzae*, some residues responsible for binding are replaced by other amino acids, thus indicating a possible loss of function or a limitation in the number of coordinated metals.

ii) PHP domain in E. coli

The PHP domain of *E. coli* DNA Pol III α subunit comprises the first 270 amino acids of the N-terminus (Fig. 1.15), close to the Thumb domain. Its structure consists of a $\alpha_7\beta_7$ TIM barrel, with seven β-strands surrounded by seven α-helices; however, there is a significant difference with regards to the structure of the other known barrels, due to the presence of the fourth β strand that runs in an anti-parallel sense, respect to the other strands, conferring a distorted disposition (Aravind & Koonin, 1998). Over the domain, there is a large loop (residues 203-240) that interacts with the Thumb domain. This loop thus forms a cover over the cavity of the barrel in which it is possible to note the presence of a phosphate molecule.
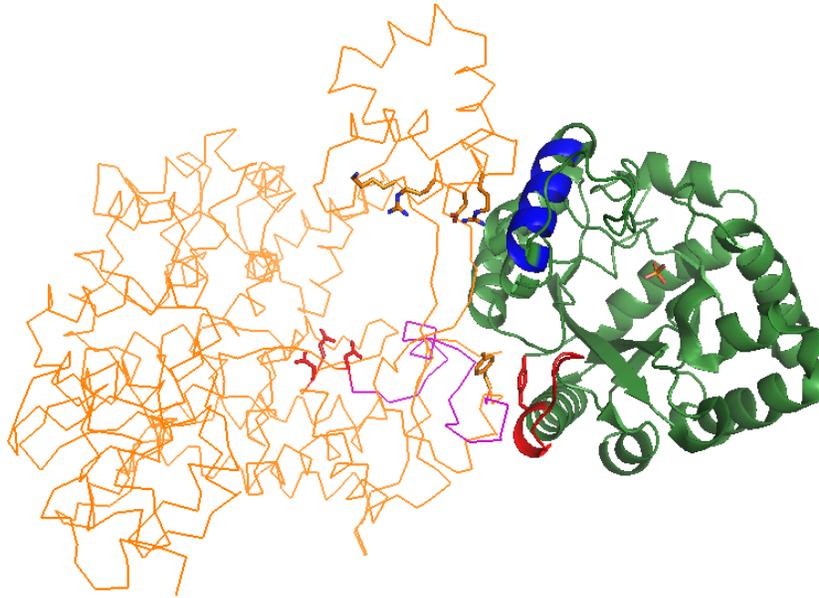
Figure 1.15: Structure of α subunit (PDB: 2HQA) with the PHP domain in green. The loop including residues 107-116 contacts the region between the Thumb and the Palm domains and is highlighted in red. The catalytic triad is shown in red whereas the binding region is in purple. The α-helix (residues 211-221) that contacts the Thumb domain is evidenced in blue. Also the charged residues of the Thumb domain that interact with the PHP domain are shown.

Furthermore, the loop including residues 203-240, together with another loop (residues 107-116), forms a groove that runs from the domain of the palm to the cavity of the barrel, constituting a space for the DNA to travel from the polymerase active site to the putative active site of the PHP domain (Lamers *et al.,* 2006). Therefore, the hypothesized pyrophosphatase activity of the latter could hydrolyze the pyrophosphate released during the incorporation of the incoming nucleotide. Actually, a phosphate molecule was found inside the pocket by crystallographic studies, representing a product of the enzymatic reaction (Lamers *et al.,* 2006). However, as explained latter, the catalytic activity of the PHP domain of α subunit is still unknown and Lamers (2006) did not found pyrophosphatase activity. The PHP domain also interacts with ε subunit, showing high conservation of those residues responsible for that contact. However, the composition of its active site is different among bacterial polymerases. In particular, the internal cluster differs in six substitutions within the nine residues responsible for binding to metal ions (Aravind & Koonin, 1998; Timinskas *et al.,* 2014) (Fig. 1.16).

Figure 1.16: Sequence representation of metal binding residues conservation in the PHP domain of Proteobacteria, according to *E. coli* DnaE positions 10, 12, 19, 44, 69, 83, 134, 201 and 203 (Timinskas *et al.,* 2014).

This divergence, associated with the loss of residues from the metal coordination cluster, suggests that the PHP domain of *E. coli* α polymerase may have developed a secondary activity not related to the exonuclease activity, which is instead present in polymerases from thermophile microorganisms. Actually, it has recently been introduced the hypothesis that the *E. coli* PHP domain has lost the site of metal coordination related to the exonuclease activity, carried out exclusively by the DNA Pol III ε subunit (Barros *et al.,* 2013).

Although the specific function of the PHP domain of α subunit of *E. coli* is still unclear, site-directed mutagenesis studies have shown that the polymerase activity can be affected by mutations in the PHP domain. For example, G134R, G133D and T260I variants suggest that the proper conduction of the DnaE polymerase mechanism is dependent on the PHP integrity (Sevastopoulos & Glaser 1977; Bierne *et al.,* 1997). The PHP NΔ60 and NΔ240 mutants, N-terminal region deleted, showed no polymerase activity and both were completely unable to interact with ε subunit (Kim & McHenry, 1996a).

Through a series of mutations in this region, it has been proven as the involvement of the PHP domain in the polymerase activity is not strictly owing to the bond established with ε subunit. In fact, the D43E mutant retains the level of polymerase activity of the *wild type* protein, showing, however, a weak affinity for ε subunit. In contrast, the D43A variant, not only underwent a minor interaction affinity between α and ε subunits, but also a clear decrease in the polymerase activity was reported (Wieczorek & McHenry, 2006). Nevertheless, in the same experience it was shown

that the ability of binding ε and the influence on the polymerase activity were not closely related among them, due to the fact that various substitutions of this residue affected both deficiencies or only one in an independent way, without significant consequences on the other (Wieczorek & McHenry, 2006). It is noteworthy that the D43 residue is conserved in all identified PHP domain sequences, although D43 is not located within the pocket of the active site and, therefore, is unlikely to have a coordination role of metal cofactors. Besides, the results of that study indicated an essential key role of the carboxyl group of the D43 residue for the correct folding and the proper enzyme activity exerted by the PHP domain and by the α polymerase.

The comparison of the tertiary structure of the PHP domain of *E. coli* DnaE protein with the PHP domain from other proteins, such as the YcdX protein of *E. coli* (Teplyakov *et al.,* 2003) or the p30 subunit of ribonuclease P of *Pirococcus horikoshii* (Takagi *et al.,* 2004), revealed the presence of the same cavity that corresponds to the active site, although with different residues configuration. The properties of the active site of the PHP domain of DNA Pol III seem to be much more similar to the DHH (aspartate, histidine, histidine) superfamily. In the bottom of the cavity of the DNA Pol III PHP domain, three aspartates (D69, D169 and D201) and two histidines (H12 and H83) were found, being necessary for metal binding; other positively charged amino acids are placed on the opposite side. This residue configuration is similar to the arrangement of residues observed in the active site of the pyrophosphatase II of *Streptococcus mutans* (Lamers *et al.,* 2006; Merckel *et al.,* 2007) (Fig. 1.17), which belongs to the DHH superfamily. The mechanism of action of the enzymes belonging to the pyrophosphatase II family is based on the coordination by three divalent metal ions, such as Mn^{2+}, Zn^{2+}, Mg^{2+} and Co^{2+}, which can occupy the binding sites and promoting the reaction with different catalytic effectiveness (Fabrichniy *et al.,* 2004). Despite the lack of confirmation about the type of catalytic activity developed by the PHP domain, which seems to be homologous to a type II pyrophosphatase, Conte (PhD Thesis, University of Bologna, 2012) demonstrated in a previous Thesis the notable pyrophosphatase activity of the isolated PHP domain of *E. coli* α polymerase (Fig. 4.3, chapter 4).
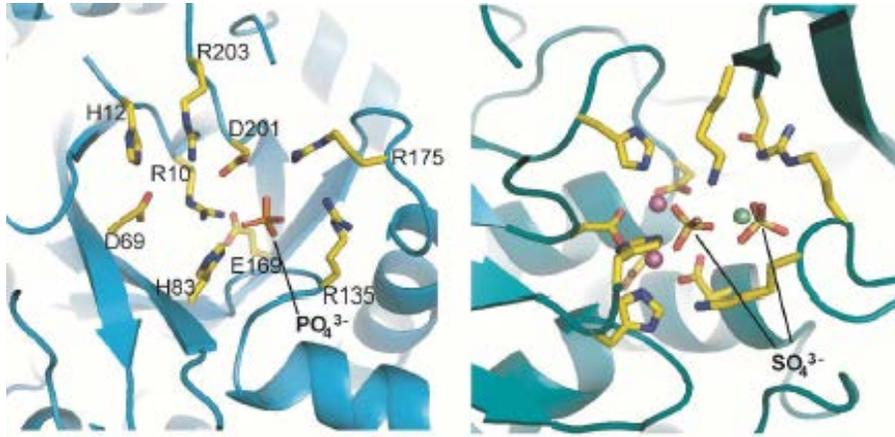
Figure 1.17: Comparison of the structures of the phosphate pocket in the PHP domain of *E. coli* α subunit (left) and the active site of type II pyrophosphatase of *S. mutans* (right). Green spheres represent Mg^{+2} ions whereas purple spheres represent Mn^{+2} ions (Lamers *et al.,* 2006).

iii) Similarities between the PHP domain and type II pyrophosphatases

Families I and II pyrophosphatases (PPases) differ in structural considerations and metal coordination. Family I PPases features only one domain, whereas family II PPases presents two domains with the active site between them (Merckel *et al.,* 2001). Nevertheless, the active site residues disposition is similar in both pyrophosphatase families. The absence or presence of substrate or product in the active site may modify the disposition and orientation of the two domains present in type II PPases: an open conformation is preferred in the absence of product or substrate, whereas the domains adopt a more closed disposition in the presence of product (Fabrichniy *et al.,* 2004; Oksanen, 2009) (Fig. 1.18).

Figure 1.18: Superposition of the N-terminal domains in an open (blue) and closed (red) conformations of *B. subtilis* PPase. Closed structure active site metals are represented by green spheres (Oksanen, 2009).
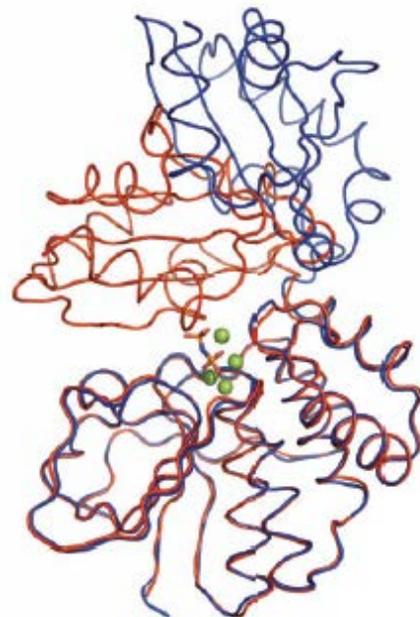
The similarity of the active site of the PHP domain in *E. coli* with the catalytic site of type II PPases has been already highlighted (Lamers *et al.,* 2006; fig. 1.17), since there is a coordination site for Mn^{+2} ,instead of an aspartate-formed Mg^{+2} coordination site as in type I pyrophosphatases. In the active site of type II PPases there are three metal ions coordinated by two histidines and four aspartates; furthermore, the pyrophosphate is placed in the inside pocket of the enzyme, interacting with two lysines, one arginine and one histidine (Fabrichniy *et al.,* 2007). Moreover, observing the structure of α subunit, it was also distinguished the presence of a molecule of orthophosphate into the PHP pocket (Barros *et al.,* 2013), which suggests the existence of a potential interaction site for phosphates in this domain.

Comparing the inner region of PHP domain with the active site of type II pyrophosphatases (Fig. 1.19), its population of amino acids could be divided in two different groups:

- Metal coordination residues (H12, S16, D19, D43, H213 and Y234);
- The remaining members of the substrate positioning (R10, H83, D201 and R203).
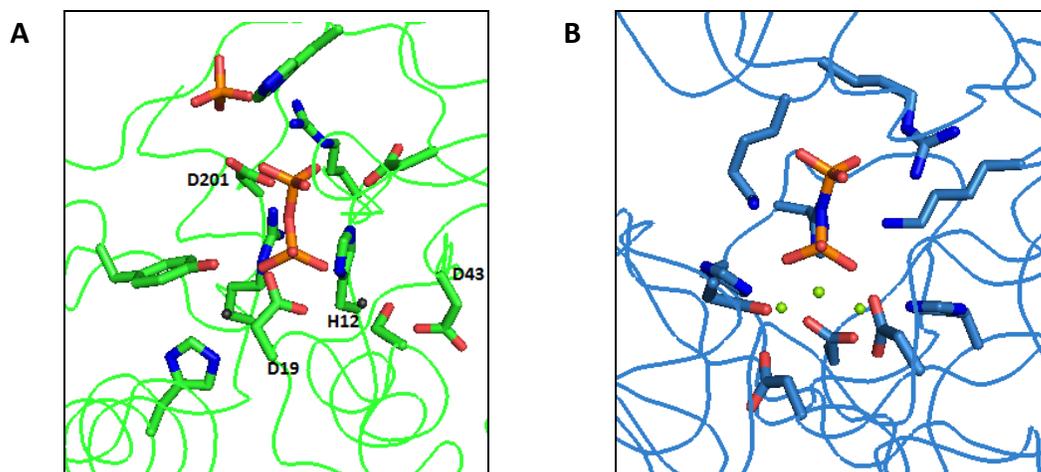
Figure 1.19: (A) Structure of the putative active site of the PHP domain of α subunit of *E. coli* (PDB: 2HNH). The evidenced amino acids, starting from the upper part, are: H83, R203, D69, D201, R10, H12, Y234, D43, D19 and H213. A pyrophosphate molecule (orange) and two metal ions (dark spheres) have been also included. **(B)** Structure of the active site of type II PPase of *B. subtilis* (PDB: 2HAW). The evidenced amino acids, starting from the upper part, are: R296, K295, K205, H9, D149, H97, D13, D75 and D15.

In agreement with the active site of type II PPases, the group of metal coordination residues of the PHP domain consists of two histidines and four negatively charged residues, while the group of adjacent amino acids, in charge of substrate interaction, contains two arginines, one aspartate and one histidine. Metal activation of Family II PPases is more complex than in Family I, requiring Mn^{+2} as activator, instead of Mg^{+2}, essential cofactor in Family I PPases. While Mn^{2+} binding affinity is constricted in the nanomolar range, Mg^{2+} binding affinity is observed in the micromolar range (Parfenyev *et al.,* 2001). The higher affinity of Family II PPases for Mn^{2+} over Mg^{2+} is partially because of the presence of histidines, instead of aspartates, as metal-binding residues: the Mg^{2+} ion prefers water or carboxylates as oxygen nucleophiles, whereas Mn^{2+} prefers softer nucleophiles as the imidazole group of histidine (Bock *et al.,* 1999).

Other ideas could be taken into consideration for the PHP domain of α subunit by comparing its structures with other well-known pyrophosphatases active site, e.g. that one from yeast pyrophosphatase (Heikinheimo *et al.,* 2001), in which two conformations (UP and DOWN) were found depending on the P2 position in the active site (Fig. 1.20). The UP conformation appears before the pyrophosphatase reaction, while DOWN conformation is present after the reaction occurs. The pyrophosphate, which bears to two metal ions (M3 and M4), reaches the active site of the enzyme. Its insertion enables the P2 phosphate to approach to metal ions M1 and M2, already present in the active site. These M1 and M2 are coordinated to Y93, Y192 and R78. Y93 coordinates P2, whereas Y192 and R78 coordinate the P1 phosphate. M1 and M2 metals promote the activation of a water molecule, thanks to the aspartate D117.

Therefore, the activated nucleophile attacks the P2 phosphate cause a movement of aspartate D117, which reaches a distance equal to 2.55 Å from the phosphate, allowing it to establish a *Low Barrier Hydrogen Bond* (LBHB). There is no longer a bridge to the hydrogen, but a stronger interaction with a shared proton between the two molecules. Consequently, there is a shift of charges on the pyrophosphate molecule that weakens, even more, the phosphoesteric bond.
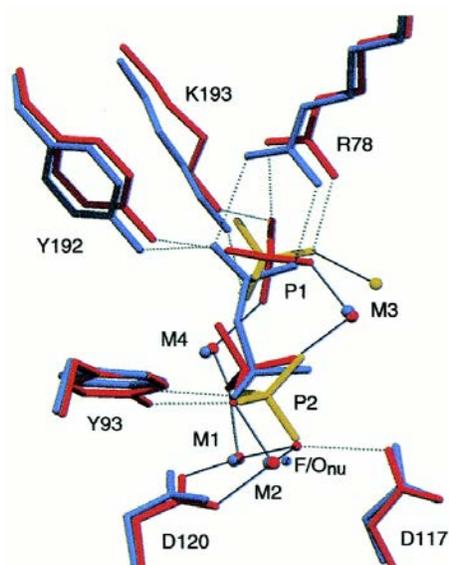
Figure 1.20: Superimposed conformations of yeast pyrophosphatase active site: with inhibitor (in blue), DOWN (in yellow) and UP (in red). UP conformation simulates substrate binding, although hydrolyzed P1 is displaced (Heikinheimo *et al.,* 2001).

At this point, there is the structural change from UP to DOWN enzyme conformation, in which aspartate D117 is close to P2 phosphate. The M3 ion moves from a position coordinated to two phosphates (P1 and P2) and the glutamate E58 to a position where it is just coordinated to P1 phosphate. The end of the reaction is characterized by the release of the P1 group, thanks to the arginine R78 and the lysine K193, whereas the P2 phosphate is stabilized by the formation of the LBHB with the aspartate D117, the tyrosine Y93 and the lysine K56.

Since the conformation of the active site of yeast inorganic pyrophosphatase is rather similar to the PHP domain structure of *E. coli*, it seems that the pyrophosphatase reaction mechanism, above described, could be performed following an analogous pattern. In the PHP structure of *E. coli*, a phosphate molecule is observed in one side of the active site, with the aspartate D201 side chain pointing towards the phosphate and toward the hypothetical metal coordination site. D201 interacts with the negative charges of phosphate at the interface of the solvent, featuring a role probably similar to that of played by the D117 of yeast inorganic pyrophosphatase, responsible for water molecule activation. Furthermore, the H12 and D19 residues, whose side chain is placed in the hypothetical P1 phosphate direction, are presumed to be involved in coordinating metal ions.

1.4.1.2 ε subunit

The ε subunit (27.5 kDa) is a member of the superfamily of DEDD DNase, classified according to the conserved DEDD domain, responsible for the preservation of the exonuclease activity (Zuo *et al.,* 2001). It is a protein composed of 243 amino acids, encoded by the *dnaQ* gene and organized in two functional domains connected by a hinge of four glutamines (Ozawa *et al.,* 2008). The catalytic domain carries out the proofreading 3' → 5' exonuclease activity; on the other hand, through the connecting region, called Q-linker, ε subunit arranges for the adequate orientation with regards to α subunit with a significant flexibility, allowing the correct coordinated performance of their respective activities (Bressanin *et al.,* 2009; Ozawa *et al.,* 2013) (Fig. 4.16 in chapter 4).

The proofreading activity of ε subunit is really useful for the correction of misincorporations introduced by α subunit during DNA replication, enhancing the accuracy of the DNA polymerase. Measurements of spontaneous mutation rates suggest that the average frequency of errors in both prokaryotes and eukaryotes is between 10^{-8} and 10^{-11} (Drake, 1969). Thus, the role of ε subunit consists of reducing the frequency of mutations by removing the mismatches, preferentially by hydrolyzing single-stranded DNA in the 3' → 5' direction, during the replication of the genome by the α subunit. Actually, mutations in the encoding *dnaQ* gene lead to the appearance of *dnaQ* mutator phenotypes, in some cases up to the point of not being viable if it is not associated with antimutator alleles of α subunit (Fijalkowska & Schaaper, 1996). The most mutagenic non-lethal variants of ε subunit are characterized by a frequency of mutations of one error every 10^4-10^5 replicated base pairs, exceeding the correction ability of the post-replicative repair systems, saturating them and leading to a high accumulation of mutations (Taft-Benz & Schaaper, 1998).

A fragment of at least three nucleotides is necessary as usable substrate by ε subunit (Miller & Perrino, 1996). The catalytic residues D12, E14, D103, D167 and H162 are located within three conserved motifs: EXO I (residues 8-21), EXO II (95-108) and EXO IIIε (128-192). These motifs are common to many proofreading subunits from the C family of polymerases. Moreover, they could be found also in other exonucleases, associated or not with polymerases, but in this case the EXO IIIε motif is, however, replaced by the EXO III motif (Blanco *et al.,* 1992).

A structural analysis carried out in a recent paper (Toste *et al.,* 2013) has highlighted the sites of interaction between α and ε subunits with the β clamp, revealing the close interaction between the C-terminal tail of ε subunit and the PHP domain of α polymerase. This location stabilizes the interaction between the polymerase and the β clamp and increases the processivity of α polymerase (Dohrmann & McHenry, 2005). The conformation of the first 186 residues of the N-terminal portion of ε subunit (ε186), containing the active site and the binding site for θ (Perrino *et al.,* 1999), was determined by both NMR (DeRose *et al.,* 2002) and X-Ray diffractometry. The last one led to a crystal of the protein complexed with two atoms of Mn^{2+} and a molecule of thymidine 5'-monophosphate (TMP), product of the catalytic reaction (Hamdan *et al.,* 2002a) (Fig. 1.21).
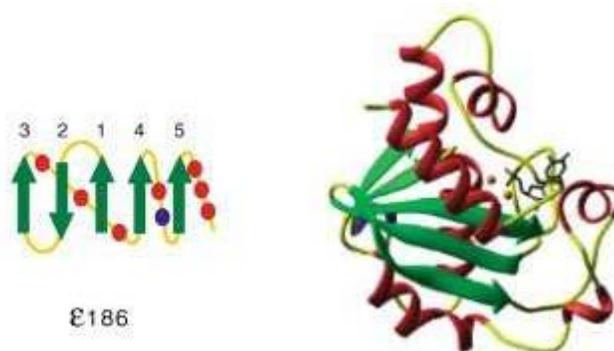
Figure 1.21: Schematic representation of the structure of ε186 in the presence of Mn^{+2} and TMP (Hamdan *et al.,* 2002a).

According to these studies, the N-terminal domain of ε is organized into six α helices that wrap a five parallel filaments-β sheet, with the exception of the β2 filament which is antiparallel (Fig. 1.21). The active site contains the catalytic residues Asp12, Glu14 and Asp167, being located in a pocket limited by the β1 strand and the α3 and α7 helices, and coordinating the two metal cofactors involved in the hydrolysis of the bond between the two nucleotides. Although the nature of the divalent ions used has not been totally determined, *in vitro* studies have shown that the protein can exploit both Mg^{2+} and Mn^{2+}, being the second one which provides a faster reaction (Hamdan *et al.,* 2002b). Finally, the N-terminal region of ε subunit forms a tight complex with θ, another protein from the DNA Pol III core, which does not have a defined catalytic activity but stabilizes ε subunit, stimulating its activity 2.5 times, approximately (Perrino *et al.,* 1999; Keniry *et al.,* 2006).

Other functions have been described for ε subunit, as its role played in the SOS response upon exposition to some mutagenic compounds. The expression of the *dnaQ* gene is in fact induced following the exposure of *E. coli* to methylmethanesulfonate (Quiñones *et al.,* 1989), while mutations in itself lead to an incomplete activation of the SOS system after exposition to nalidixic acid (Pohlhaus *et al.,* 2008). ε subunit is also probably involved in the translesion synthesis (TLS), performed by DNA polymerases IV and V, being able to physically interact with DNA Pol V UmuD and UmuD' subunits (Sutton *et al.,* 1999) at the C-terminal region of ε subunit. It has also been reported in literature that an increase of the supercoiling DNA can lead to an increased expression of the *dnaQ* gene (Quiñones & Neumann, 1997). This correlation is interesting from a physiological point of view because the topological status of DNA varies depending on the environmental conditions, the cell phase or following a prolonged scarcity of resources; therefore, the existence of a link between these factors and the level of expression of ε subunit allows us to hypothesize that the presence of a more or less marked proofreading activity has a significant role in adaptive mutagenesis in *Escherichia coli*.

1.4.1.3 θ *subunit*

The θ subunit is the smallest (8.6 kDa) protein component of the DNA Pol III, and its catalytic activity has not been defined yet. Its encoding gene (*holE*), whose expression level is low (Carter *et al.,* 1993), can be deleted from the chromosome (ΔholE) without producing a detectable phenotype or increasing the mutation rate (Slater *et al.,* 1994). Therefore, it could be interpreted that θ is not essential for DNA Pol III core functions and that its deletion does not affect the synthesis of DNA. Nevertheless, several studies reflect the important role of θ in stabilizing the DNA Pol III core. Furthermore, there is a close interaction between θ and ε, although not between θ and α subunits (Studwell-Vaughan & O'Donnell, 1993; Jonczyk *et al.,* 1998).

The θ subunit is able to increase about 2.5 times the 3' → 5' exonuclease activity of ε subunit (Studwell-Vaughan & O'Donnell, 1993). Hence, θ, although not essential, may indirectly play a key role in DNA replication, reducing the mutation frequency through its interaction with the N-terminal region of ε subunit (Perrino *et al.,* 1999).

Recent studies (Taft -Benz & Schaaper, 2004) showed that θ could be a stabilizing factor for ε, which is inherently unstable (Foster & Marinus, 1992), being also able to raise the temperature of thermal inactivation of ε about 14° C (Hamdan *et al.,* 2002b). It has been reported that the association of θ with ε-186 also decreases the formation of aggregates in mixtures of organic solvents and water (Gupta *et al.,* 2004), improving the purification process. In addition to the stabilizing effect on ε subunit, θ could take the free ε subunit form in the cell and place it in the catalytic core; in this case, ε should be much more stable than the non-attached form (De Rose *et al.,* 2003). Moreover, the capacity of binding between α and ε increases in the presence of θ (Taft - Benz & Shaaper, 2004). Therefore, θ could be essential in the assembly and the stabilization of the DNA Pol III catalytic core, thus improving the purification yield.

In the replicative complex of Gram negative bacteria, the polymerase and exonuclease enzymes are two separate proteins. In other organisms, in which the polymerase and exonuclease activities are part of the same polypeptide, proteins corresponding to θ subunit were not found. This suggests, once again, that θ protects the ε subunit (Taft -Benz & Schaaper, 2004).

A protein from bacteriophage P1, which is able to replace θ in the core of the Pol III, was already identified and it was named HOT (Homolog Of Theta) protein after its characteristic function (Lobocka *et al.,* 2004). The HOT protein features a degree of homology equal to 65% and a degree of identity of 53%, when compared to θ subunit. The replication of the P1 phage depends almost entirely on the host replication system, since this phage does not code for its own polymerase. Therefore, the presence of a encoding gene for a homologue of θ in a compact, limited genome, assigns HOT, and consequently θ, a certainly important role, although not identified yet. Since ε is unstable, HOT protein may act as a stabilizing agent to provide a benefit to the phage, which presumably has available a greater amount of core (Chikova & Shaaper, 2005). The presence of HOT neither complements nor changes the frequency of mutations resulting from *holE* deletion, but changes their specificity: in the case of the deleted strain, AT $\rightarrow$ TA transversions predominates, while a HOT expressing strain features a greater frequency of GC$\rightarrow$ AT transitions, showing that the two proteins act in a different manner (Chikova *&* Shaaper, 2005).
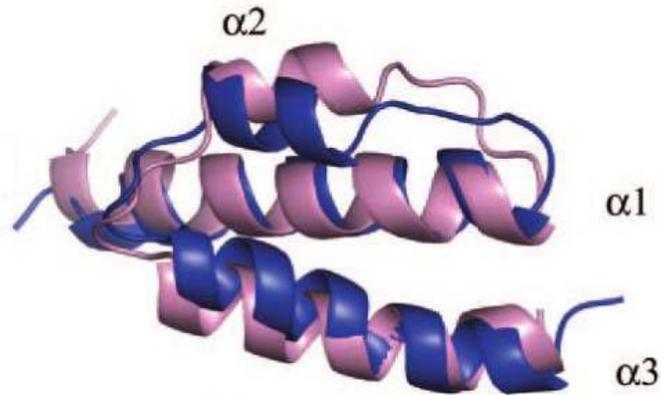
Figure 1.22: Structure alignments of θ subunit (in blue) and HOT protein (in violet) (Mueller *et al*., 2005).

The structure of θ was determined by NMR in water-ethanol solution (Mueller *et al.,* 2005) (Fig. 1.22). The protein is organized in three α-helices, in the α1-α2-α3 order, and presents the N-terminal region containing hydrophobic residues (Keniry *et al.,* 2006), which made difficult the data collection in water. Structural information about the complex formed with ε186 was achieved by NMR techniques (Kirby *et al.,* 2006).

1.4.2 Sliding Clamp (β2)

The Sliding clamp is a β2 homodimer, formed by the link between two β-protomers, each one formed by three motifs. Inside the dimer, there is an inner channel of about 35 Å diameter, constituted by positively charged α-helices, which are able to accommodate a double-stranded DNA (Georgescu *et al.,* 2008) (Fig. 1.23). The sliding clamp plays an important role in the replication process, ensuring a stable interaction between the DNA and the DNA Pol III catalytic core. In particular, Arginine 24 and Glutamine 149 residues, exposed at the surface of the loop of the C-terminal face, make contact with the DNA phosphates and induce a tilt of about 22°, allowing the load of β-clamp onto the DNA strand (Georgescu *et al.,* 2008).

DNA loading and conformational movements of the sliding clamp are governed by the DnaX complex, denominated Loading clamp, which is able to bind β2 complex to the DNA, in close contact with the catalytic core, thereby ensuring a higher processivity. In fact, the sliding clamp cannot be assembled onto the DNA autonomously, due to the fact that it requires the intervention

of the DnaX complex that uploads and blocks the β clamp onto the template. Although the core is sufficient to replicate a DNA strand in the presence of a primer, even in the absence of other subunits of the holoenzyme, binding to β clamp is essential to achieve the proper replication of the entire chromosome. The isolated αεθ shows a rate of polymerization about 20 bp/s. However, the association with the β2 dimer allows the achievement of higher values of replication speed, up to 1000 bp/s (Jay *et al.,* 1981). Hence, the β clamp represents the most important processivity factor of the DNA polymerase III.
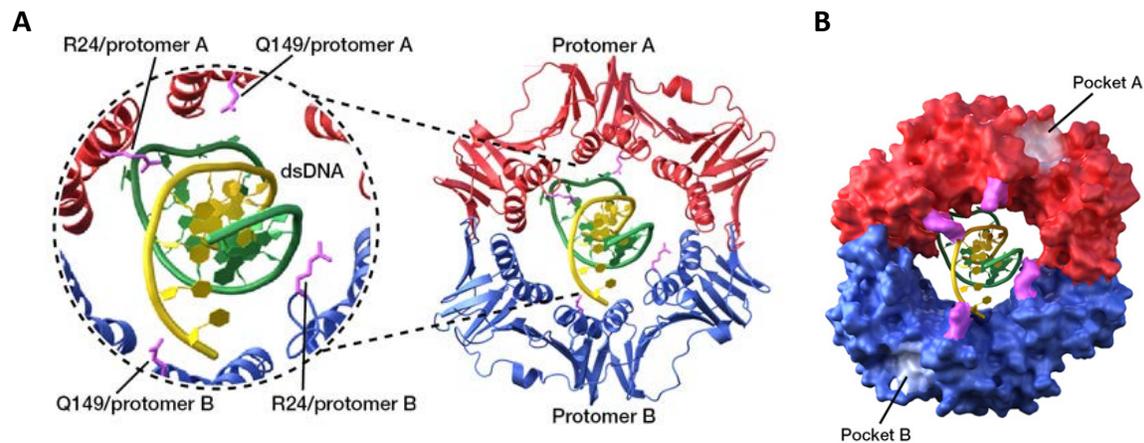
Figure 1.23: (A) Structure of β2 dimer bound to DNA. R24 and Q149 residues are highlighted in pink. **(B)** Hydrophobic pockets A and B, placed between domains II and III, pointing towards the link with α subunit (Georgescu *et al.,* 2008).

The β2 dimer features two hydrophobic pockets: the first one, located between the domains II and III of a β-monomer, binds to the α subunit; the other one remains free in order to establish interactions with other polymerases (Fig. 1.23.B) (Naktinis *et al.,* 1995). Under conditions that cause DNA Pol III stalling during DNA replication, the sliding clamp allows the polymerase switching, in which DNA polymerase IV have access to the DNA during the course of the new strand synthesis after Pol III stalling (Indiani *et al.,* 2005). The area of DNA binding between H148 and R152 residues of the C-terminal region is essential for maintaining the holoenzyme functionality and is also implicated in the exchange of polymerases during the replication of damaged DNA strands. Furthermore, studies with β sliding clamp mutants have shown that DNA polymerases II and IV, but not DNA Pol III, compete with β for DNA binding as they interact with the same region (Heltzel *et al.,* 2009b).

The β clamp also interacts with components of the DnaX complex (López de Saro *et al.,* 2003), the DnaB helicase, MutS and MutL proteins (López de Saro *et al.,* 2006), the cell regulatory Hda factor (Kurz *et al.,* 2004) and also with DNA Pol I at the end of the synthesis of an Okazaki fragment (López de Saro *et al.,* 2003).

1.4.3 Loading Clamp (τγδδ'χψ)

The Loading Clamp is an oligomer consisting of τ (71.1 kDa) and γ (47.5 kDa) subunits, both encoded by the *dnaX* gene, in association with δ (38.7 kDa), δ' (36.9 kDa), χ (16.6 kDa) and ψ (15.2 kDa) subunits, encoded by *holA, holB, holC* and *holD* genes, respectively. The different proteins are assembled together in a pentameric ring, consisting of three DnaX monomers (τ or γ), associated with individual δ and δ' subunits. ψ and χ subunits are also linked to the DnaX complex, thanks to the interaction established by ψ (Dallmann, 1995).
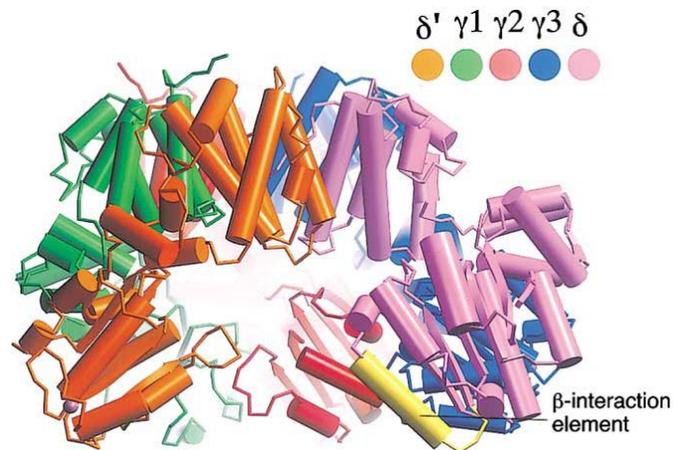
Figure 1.24: Structure of the Loading clamp complex with three γ subunits in the absence of τ subunit (Jeruzalmi *et al.,* 2001a).

Both τ and γ subunits are products of *dnaX* gene translation. During this process, a ribosomal frameshift (Fig. 1.25) is generated with a frequency equal to 40%, producing a premature stop codon. This event causes the displacement of the frame, arresting the *dnaX* gene translation so as to create the truncated γ protein (Tsuchihashi & Kornberg, 1990). The τ subunit consists of five domains, although only the first three domains (I-III) are responsible for ATPase activity and oligomerization. The IV domain includes the binding site to the DnaB helicase (Gao & McHenry, 2001a), whereas the V domain features binding sites to α subunit (Gao & McHenry, 2001b).

The first three domains of τ subunit, responsible for ATPase activity and oligomerization, are shared with γ subunit (Fig. 1.25); consequently, γ subunit is identical to the 431 residues of the N-terminal fragment of τ subunit. Anyway, genetic studies indicate that τ, but not γ, is essential for bacterium viability (Blinkova *et al.,* 1993) because of its unique interactions with α and DnaB to constitute a functional replisome. During DNA replication, the DNA Pol III HE requires, at least, two cores assembled with the C-terminal portion of two τ subunits. The γ protein, lacking this region, cannot guarantee the binding to the polymerase core; therefore, in proposed models, DnaX complex is composed of at least two τ subunits.
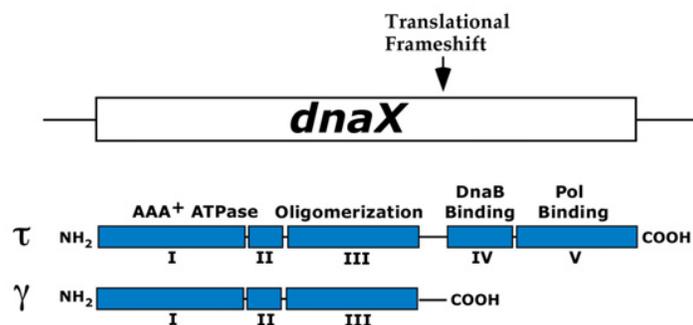
Figure 1.25: Representation of the *dnaX* gene and the domains of τ and γ subunits *(McIrnery et al., 2007).*

In the DnaX complex, δ' and ψ subunits bind directly to γ, whereas δ and χ bind to δ' and ψ, respectively (Onrust *et al.,* 1995). The use of cross-linking techniques has also highlighted that δ' and ψ subunits bind exclusively to γ and not to τ (Glover & McHenry, 2000). The structure of the γ3δδ' complex has been crystallographically resolved, and appears as an opened-circular ring with δ and δ' subunits located at the two extremes (Jeruzalmi *et al.,* 2001a). The tertiary structure of δ subunit linked to β is also available (Jeruzalmi *et al.,* 2001b).

The complex is responsible for contacting the primosome, the Pol III αεθ and the β-clamp, ensuring the cohesion of close proteins near the replication fork and thus directing the replication process. Both τ and γ subunits possess ATPase activity that provides essential energy for conformational changes of the DnaX complex, controlling the assembly with DNA and the coordination of αεθ movements. Actually, through the interaction of τ or γ subunits with ATP, a conformational change of the entire complex allows δ subunit to make contact with the sliding clamp, which in turn assumes a suitable conformation for DNA positioning onto β2 dimer (Fig.

1.26). The settling of the β clamp on DNA stimulates ATP hydrolysis and the release of the DnaX complex from the β2 clamp (Jeruzalmi *et al.,* 2001a). The correct interaction of the sliding clamp and the catalytic αεθ with DNA is achieved by the interaction that χ subunit performs with the SSB proteins (Witte *et al.,* 2003).
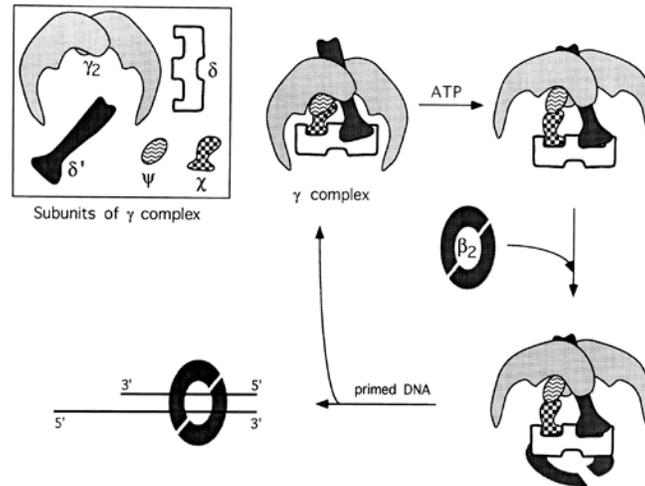
Figure 1.26: Cartoon representing the activation of the γ complex by ATP. The presence of ATP induces a conformational change that enables δ subunit to bind to the β sliding clamp, which is subsequently transferred onto primed DNA (Naktinis *et al.,* 1995).

The function of δ and δ' is to facilitate and coordinate the opening of the β clamp during the phases of DNA association and dissociation. In particular, δ binds to the β dimer, destabilizing the interface with DNA and thus inducing the dissociation (Naktinis *et al.,*1995; Jeruzalmi *et al.,* 2001b), confirming that β releases from DNA as a result of ring opening. Therefore, δ subunit is considered as a "ring opener" enabling β to slip from DNA (Turner *et al.,* 1999). In contrast, δ' may serves as a "clamp assembly modulator", facilitating the interaction of δ with β for ring opening, and then by promoting the dissociation of δ from β, closing the ring around DNA (Turner *et al.,* 1999). The function of χ and ψ subunits has not been precisely defined yet, even though it was demonstrated that χ binds to SSB proteins, detaching the primase from DNA at the end of the synthesis of a primer, before the loading of β clamp (Yuzhakov *et al.,* 1999). On the other side, ψ subunit could be involved, as well as in binding χ to τ, in stabilizing or promoting conformational changes of the DnaX complex, induced by ATP hydrolysis (Anderson *et al.,* 2007).

The exact composition of the DNA polymerase III holoenzyme has not been established yet, due to the ambiguity and difficulty for *in vivo* examination of the precise stoichiometry of the loading clamp central pentameric assembly. The classic model stated that χ and ψ proteins are associated to the pentamer τ2γδδ'; however, other combination possibilities have been isolated for DnaX complex assembly (γ3δδ', τγ2δδ', τ2γδδ', τ3δδ') (Park *et al.,* 2010) (Fig. 1.27). The τ-only and γ-only complexes assemble preferentially over the mixed complexes.
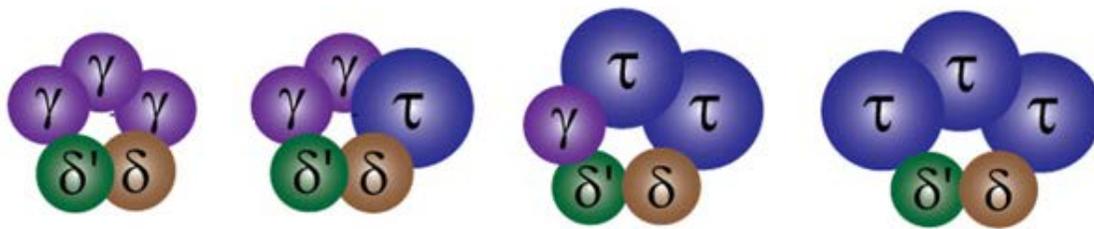
Figure 1.27: Different states of oligomerization of the pentameric DnaX complex, *in vitro* assembled (Park *et al.,* 2010).

Moreover, the characterization of an *in vitro* τ3δδ' pentameric ring assembly introduced the possibility that the replisome may act according to a trimeric model with three polymerase cores (McIrnery *et al.,* 2007) (Fig. 1.28). In this case, Pol III holoenzyme would consist of three cores arranged in such a way to provide two active cores for lagging strand polymerization, thus ensuring a synthesis rate comparable to that of leading strand, as it has been also proposed for the replicative complex of bacteriophage T7 (Jhonson *et al.,* 2007) or for the polymerase of *Sulfolobus solfataricus* (Mikheikin *et al.,* 2009). This model could explain how the polymerization mechanism proceeds in parallel and simultaneously for both strands. Therefore, some authors support the idea of a trimeric sub-assembly over the classic dimeric polymerase presented by McHenry (1982). This new model would increase processivity and fidelity of replication as well as a more efficient lagging strand synthesis. Georgescu and co-workers (2013) highlighted that a holoenzyme containing three cores is actually more efficient in the *in vitro* synthesis of Okazaki fragments. In addition, fluorescence analysis of individual molecules have shown a regular *in vivo* replacement of catalytic cores on the lagging strand (Lia *et al.,* 2012), suggesting that the synthesis of Okazaki fragments may be carried out through the coordinated activity of more polymerases. However, a trimeric sub-assembly of the DNA Pol III has not been obtained *in vivo*.
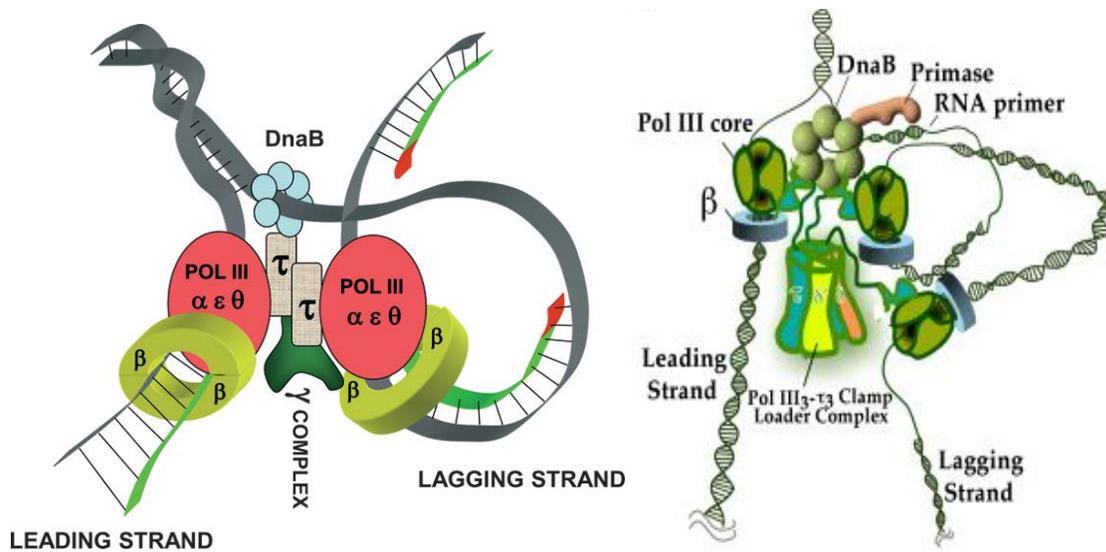
Figure 1.28: Comparison of the classic dimeric model (left) with 2 τ subunits and 2 catalytic cores, proposed by McHenry (1982) (Figure from Fijalkowska *et al.,* 2012); and the new trimeric model (right), proposed by McIrnery (2007) for *E. coli* DNA Pol III, with 3 τ subunits and 3 αεθ (McIrnery *et al.,* 2007).

1.5 DNA polymerase IV

DNA Pol IV is one of the most relevant translesion-synthesis (TLS) polymerases of *E. coli*, being part of the Y family polymerases (Wagner *et al.,* 1999). DNA Pol IV is encoded by the *dinB* gene, characterized for being a damage-inducible gene, and its intracellular protein concentration is high (250 molecules/ cell), increasing even 10-fold after SOS induction (Kim *et al.,* 2001). Unlike other polymerases, DNA Pol IV lacks intrinsic exonuclease activity and, consequently, its fidelity is expected to be low (Goodman, 2002). Actually, it has been observed that Pol IV is able to generate frameshift mutations in stationary phase populations (Tompkins *et al.,* 2003). Nevertheless, Pol IV does not contribute significantly to the frequency of mutations in growth phase cells and its deletion causes a reduction of 5-10 times the number of adaptive mutants (Hersh *et al.,* 2004). DNA Pol IV also overcomes the replicative block of the DNA Pol III (Goodman, 2002).

During SOS induction, Pol IV production increases 10-fold (Kim *et al.,* 2001) and one of the exerted functions is to interfere with DNA Pol III, creating a checkpoint for stopping replication and allowing the repair of DNA lesions via the suitable repair pathway. This function is crucial in eukaryotic cells and could be the reason for which, in all organisms investigated, there are enzymes homologous to the *E. coli* Pol IV (Ohmori *et al.,* 2001). In contrast, the DNA Pol V is only present in prokaryotic organisms (Ohmori *et al.,* 2001).

DNA Pol IV is also responsible for performing translesion synthesis (TLS) at the replication fork, e.g. bypassing N2-deoxyguanine adducts (Jarosz *et al.,* 2007). Some studies demonstrated that when DNA Pol IV was absent in growing cells, spontaneous mutations remained practically constant (McKenzie *et al.,* 2003; Wolf *et al.,* 2004), concluding that the activity of DNA Pol IV at the replication fork is likely to be limited, at least, under normal growth conditions. On the contrary, its overproduction led to a clear mutator phenotype (Wagner & Nohmi, 2000; Kuban *et al.,* 2005).

1.6 DNA polymerase V

DNA Pol V, encoded by the *umuDC* operon, is a TLS polymerase from the Y family of polymerases (Wagner *et al.,* 1999; Tang *et al.,* 1999). The heterotrimeric DNA Pol V (UmuD'$_2$C), consists of a dimer of the post-translationally modified UmuD' protein, attached to the UmuC protein (Goodman, 2002) that represents the catalytic domain. Pol V is present in the cells approximately 45 minutes after UV treatment, working as an error-prone enzyme, as it lacks proofreading activity. Its base insertion fidelity is one of its most characteristic features, making possible to copy across a variety of DNA lesions, independently of the level of detected mutagenesis and being, therefore, a TLS polymerase (Reuven *et al.,* 1999; Tang *et al.,* 1999). There are 15 molecules of DNA Pol V per cell (Woodgate & Ennis, 1991), indicating that Pol V is hardly observable but not completely absent (Bhamre *et al.,* 2001). Due to that, DNA Pol V is unlikely to play a significant role in chromosome replication. Nevertheless, after SOS induction, the level of UmuD and UmuC increases up to 2400 and 200 molecules per cell, respectively (Kim *et al.,* 2001). This is equivalent to the basal level of DNA Pol IV without SOS activation (250 molecules per cell). Nevertheless, the level of error-prone polymerases must be strictly controlled in order to guarantee a low rate of mutations under growth conditions.
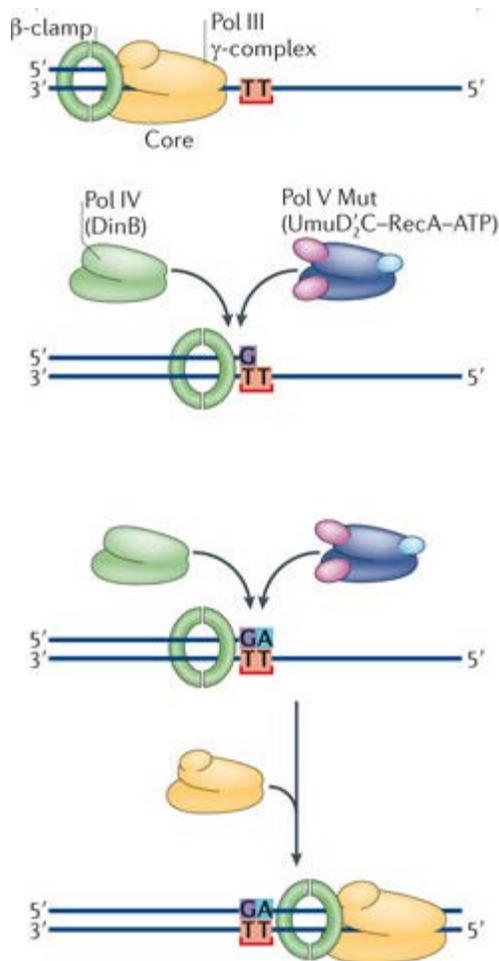
Figure 1.29: Activity of TLS polymerases (DNA Pol IV and DNA Pol V) in *E. coli* solving a DNA block during replication but maintaining some of the introduced mistmaches (Sale *et al.,* 2012).

The reason for which DNA Pol V is very scarce can be explained by examining the proteolytic regulation exerted on UmuD. This protein is degraded by the Lon protease, according to an ATP-dependent mechanism. The signal for protein degradation resides at the N-terminus of UmuD, region that is lost when the protein is processed in UmuD' (González *et al.,* 1998). Only in those cells in which DNA damage occurs, some UmuD molecules are converted from RecA into UmuD', which is relatively resistant against Lon-mediated proteolysis (Frank *et al.,* 1996).

As a consequence of the Pol IV and Pol V overexpression after SOS system activation, there is an increase of spontaneous mutagenesis, known as "SOS mutator phenotype", with the competition between DNA Pol III replicase and DNA Pol V (Fijalkowska *et al.,* 1997) for repairing mismatches introduced by the α polymerase. DNA Pol V, as well as Pol IV, lacks the 3' → 5' exonuclease activity, being thus considered as an unfaithful polymerase (Tang *et al.,* 2000). Nevertheless, in the presence of a mismatch in the newly synthesized DNA strand in which DNA Pol III holoenzyme is not able to continue DNA elongation, accessory polymerases, as DNA Pol V, should have access to the replication site. DNA pol V enables the overcoming of DNA synthesis block but, unfortunately, introducing mutations at high frequency always with the purpose of avoiding the block of an essential cell function as DNA replication (Sale *et al.,* 2012) (Fig. 1.29). SSB proteins and β-sliding clamp may participate in this phenomenon increasing the processivity of DNA Pol V (Schlacher *et al.,* 2005).

III - DNA REPLICATION

The viability of an organism depends on the successful replication of its genome, a process linked to DNA polymerases. Basically, the DNA polymerase develops the complementary incorporation of dNTPs in the 5' → 3' direction (Kornberg & Baker, 1992). Despite the diversity found among DNA polymerases, a common mechanism of action has been established to all of them (Steitz *et al.,* 1994). These enzymes carry out the synthesis of DNA via an ordered mechanism in which the enzyme binds to a DNA primer/template molecule, which serves as substrate, resulting in a polymerase-DNA binary complex.

Nucleotide incorporation is initiated by the binding of a dNTP to the enzyme-substrate complex in order to develop the DNA polymerase-DNA-dNTP ternary complex. The Fingers domain regulates the correct positioning of dNTPs at the polymerase active site (Minnick *et al.,* 1996). The conserved residues located in this domain are fundamental to interact with the two substrates of the catalytic reaction: the incoming nucleotides and primer/template DNA. In the absence of nucleotides, the Fingers domain is in the open conformation and, once bound to the nucleotide, this domain rotates towards the Palm domain, thereby closing the active site.

DNA polymerization involves the presence of divalent metal ions A and B, coordinated by two catalytic carboxylic residues, conserved both in DNA and RNA polymerases (Brautingham & Steitz, 1998). The metal A enables the nucleophilic attack to the α phosphate of the incoming dNTP, forming a phosphodiester bond. On the other hand, metal B mediates the release of the inorganic pyrophosphate molecule (PP) generated after the reaction (Fig. 1.30). Metal ions stabilize the structure and charge of the pentavalent transition state, generated during the course of the reaction, being an universally conserved mechanism among all DNA polymerase families (Joyce & Steitz, 1994). If nucleotide incorporation is wrong, the exonuclease domain, if present, controls its removal.
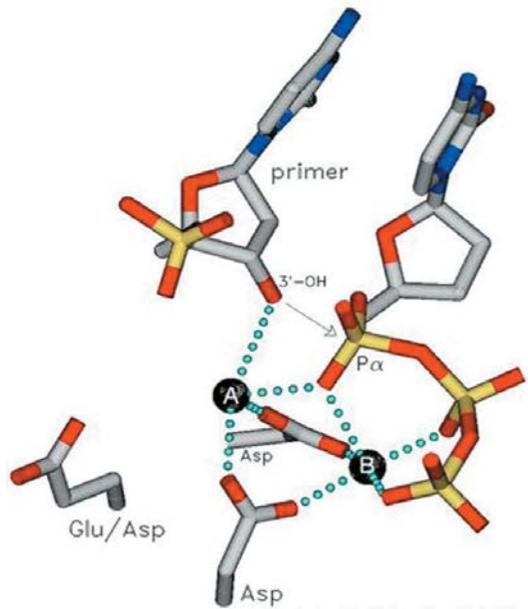
Figure 1.30: Schematic diagram for metal coordination of the dNTP incorporation mechanism of DNA polymerases. Metals A and B are shown as black spheres (Gleghorn *et al.,* 2011).

Some polymerization properties, such as nucleotide insertion efficiency (Beard *et al.,* 2002) or the number of incorporated nucleotides in each substrate binding event (Blanco *et al.,* 1989), could differ among polymerases. There are also divergences related to the substrate preference, e.g. DNA with single nucleotide gaps (Pol β) or the preferential use of damaged DNA (Pol μ) (McCulloch *et al.,* 2004).

1.7 DNA replication in E. coli

The initiation of replication is tightly coordinated with the cell cycle (Boye *et al.,* 1996), and the synthesis of a complete copy of the chromosome (4.6 Mbp) takes about 40 minutes, assuming a rate of incorporation of 1000 bp/s, approximately, when all components of the DNA Pol III HE are present (Fig. 1.31).

In *E. coli*, the duplication of the genome needs the formation of a bubble of replication at the *oriC* region, proceeding bidirectionally on both strands, until the termination (*Ter)* region, at the opposite end of the chromosome. In that bubble of replication, DNA strands are replicated in parallel by, at least, two polymerase cores. This process requires the activation of a complex mechanism of replication in which more proteins ought to be involved. The Pol III HE, in fact, interacts with other proteins forming part of the replicative complex, such as the DnaB helicase, the DnaG primase and the SSB proteins.
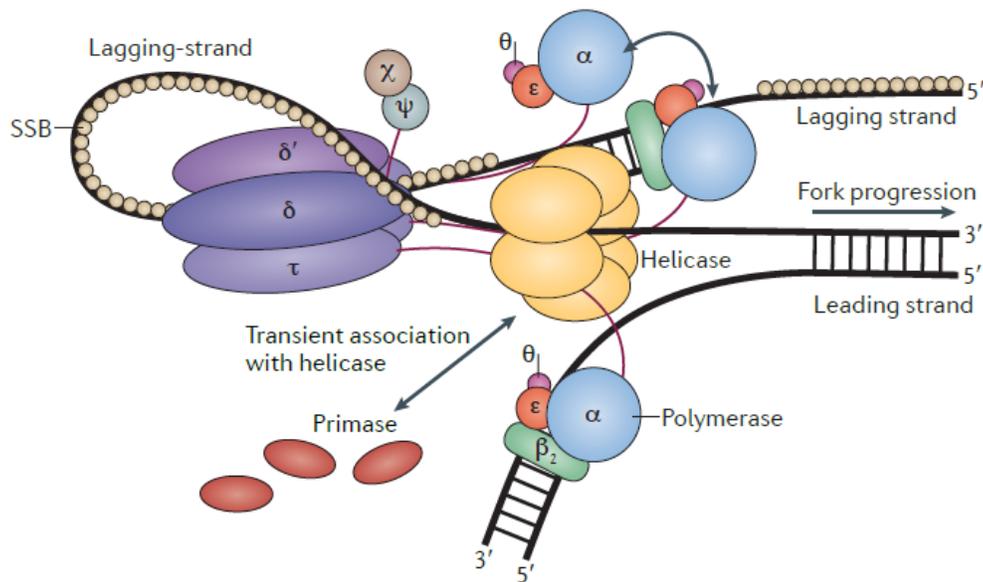
Figure 1.31: *E. coli* DNA Pol III HE representation according to the trimeric model. Also other proteins co-involved in DNA replication (SSB proteins, Primase and Helicase) are present (Robinson *et al.,* 2013).

The first stage of the replicative process is the unwinding of the double helix of DNA at the *oriC* region by the DnaB helicase, an event triggered by the denaturating action of DnaA protein on some serially repeated sequences rich in A-T pairings and promoted by the intervention of the DnaC protein, which complexes with DnaB. Once DNA denaturation is carried out, the replicative helicase (DnaB), together with DnaG primase, forms the DnaB6-DnaG3 primosome, which will accompany the polymerase replication. The hexameric helicase proceeds in the 5' → 3' sense on the lagging strand while the primase synthesizes the short RNA primers for each Okazaki fragment and one for initiating the continuous synthesis of the leading filament (Mott & Berger, 2007). The clamp loader, contacting the DnaB helicase, reaches a stable attachment of the β-clamp to the RNA primers (Kim *et al.,* 1996b).

The replication process takes place through a system of alternating polymerization, continuous for the leading strand and discontinuous for the lagging filament. The polymerization mechanism generates discontinuous segments of DNA on the lagging-strand named Okazaki fragments, of about 1000-2000 nucleotides long (Okazaki *et al.,* 1967). This fact induces a folding of the lagging strand between the replication fork and the anchor site of the polymerase, being in where the association between the single strand and the SSB proteins occurs. The single-stranded DNA is covered by SSB proteins, which protect DNA from degradation, remove secondary structures from the DNA and avoid negative interferences with other polymerases at the replication fork.

During DNA replication, DNA polymerases cycle on and off the lagging strand, and a fast dissociation process is necessary after extending each Okazaki fragment for an optimal initiation and synthesis of the following fragment (Kornberg & Baker, 1992). *E. coli* Pol III HE triggers two types of recycling mechanisms: the "collision release", in which the polymerase contacts with another downstream Okazaki fragment but immediately binds with another sliding clamp, already attached to a synthesized RNA primer (Stukenberg *et al.*, 1994). In this mechanism it has been proved that τ subunit takes part in separating the Pol III from DNA, using the energy produced by its ATPase activity to release the β clamp (Jeruzalmi *et al.,* 2001a; Lopez de Saro *et al.,* 2003); a second mechanism, which is known as "signaling release", is characterized by the release of DNA Pol III from the sliding clamp before completing the Okazaki fragment synthesis (Hamdan *et al.*, 2009). Therefore, it can be concluded that the existence of two hypotheses for polymerase-recycling systems on the lagging strand sorts out the relevance of new considerations about the coordinated synthesis of the leading and lagging strands by DNA Pol III HE.

Finally, the discontinuous synthesis of the lagging strand requires the replacement of RNA primers by DNA sequences on the newly synthesized DNA strand, so as to turn a single strand DNA into a double helix. These functions are performed by the DNA polymerase I, which removes RNA primers by its 5' → 3' exonuclease activity and replaces them with DNA fragments. The 3' extreme of different DNA fragments are welded to the next primer by the action of the DNA ligase, which results in the complete formation of the neo-synthesized strand.

1.8 Fidelity and processivity of replication

DNA replication is characterized by being highly accurate, incorporating complementary dNTPs to the template. The largest contribution to the fidelity of replication is provided by the DNA polymerase, capable of incorporating a specific nucleotide in each catalytic cycle, instead of nucleotide analogues or non-complementary nucleotides. The fidelity of the replication process is very high and features to an error every 10^9-10^{10} base pairs (Drake, 1991). There are three main factors involved in ensuring the maintenance of this fidelity (Fig. 1.32): the selection of the incoming nucleotides, exerted by α subunit according to the pairing rules (Watson & Crick, 1953); the 3' → 5' exonuclease activity of ε subunit, for removing the mismatches introduced by α subunit during replication, by hydrolysing single-stranded DNA in the 3' → 5' direction and thus reducing the frequency of mutations; and the mismatch repair exerted by MutH, MutL , MutS and MutU proteins (Schaaper, 1993).
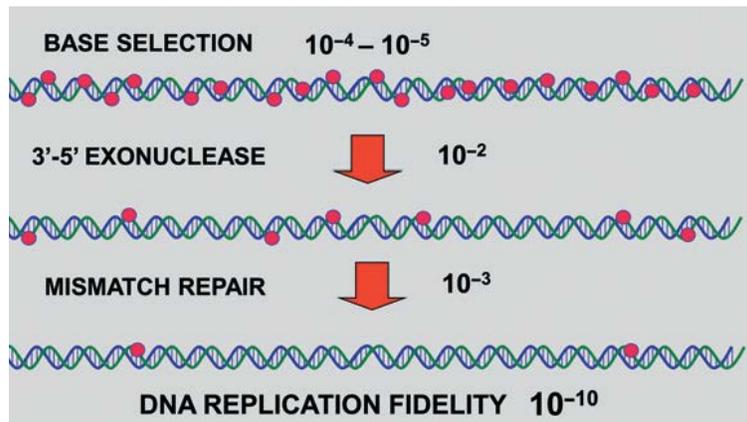
Figure 1.32: Three serial fidelity mechanisms during chromosomal replication can produce the low error rate equal to 10^{-10} errors per base per round of replication (Fijalkowska *et al.,* 2012).

When a wrong dNTP is incorporated, the addition of the next nucleotide is strongly inhibited owing to the improper orientation of the initiating chain end in the polymerase active site. DNA polymerases separate the base pairs formed, allowing the switching of the end of the initiator chain to the exonuclease active site, in order to eliminate non-correctly incorporated dNTPs.

Based on structural studies of DNA Pol I (Shamoo & Steitz, 1999), a rotation of duplex DNA was observed for this switching. This conformational rearrangement is triggered by the Thumb domain, which is in direct contact with the DNA strand. Recently, conformational changes on the primer following the addition of an incorrect nucleotide have been studied by Förster Resonance Energy Transfer (FRET) (Christian *et al.,* 2009; Berezhna *et al.,* 2012; Markiewicz *et al.,* 2012; Lamichhane *et al.,* 2013). These conformational changes must be related to the primer switching from the *pol*-site to the *exo*-site of Klenow enzyme.

One of the most important requirements in replicases is to have high processivity for the continuous, efficient, complete replication of the genome. Most of the DNA polymerases are able to incorporate only few nucleotides per DNA binding event. Usually, the procedure of accessory factors or "processivity factors" confers the necessary anchoring between DNA and DNA polymerases for the required processivity of the replication process. The sliding clamp is considered to be the most important factor in DNA Pol III processivity. Actually, its association with α subunit enables the polymerase to incorporate even 1000 nucleotides per second (Maki & Kornberg, 1988), a rate which is lowered to 20 nucleotides per second in the absence of the sliding clamp (Fay *et al.,* 1982). Furthermore, there are other minority factors, depending on the organism, which contribute to the processivity of the replication process. This is the case of thioredoxin, responsible for the high processivity reached by the DNA polymerase of bacteriophage T7 (Bedford *et al.,* 1997), despite not contacting the DNA directly.

IV - OBJECTIVES

The first objective of this Thesis was to design an efficient method to produce the αεθ complex and characterize its *in vivo* assembly. A dual-plasmid system was selected for protein co-expression, using a high-copy number plasmid for the overexpression of α subunit and a low-copy number vector for overexpressing τ, ε, and θ subunits. This method could clarify the assembly *in vivo* of *E. coli* DNA Pol III, due to the fact that τ subunit is responsible for binding the different catalytic cores of the holoenzyme through its contact with the α subunit (Blinkova *et al.,* 1993).

A second aim of this Thesis was to characterize the function and the active site of the PHP domain of α subunit, whose specific activity is still unknown. Previous in silico studies performed in my host laboratory found a structural similarity between PHP domain and type II pyrophosphatases, suggesting a possible pyrophosphatase role that would be in agreement with published structural data. Furthermore, a previous Thesis carried out in my host laboratory had highlighted that the isolated PHP domain features a clear pyrophosphatase activity (Conte, 2012). With the aim of defining the catalytic active site of the PHP domain and to evaluate the coupling between polymerase and pyrophosphatase activities, both carried out by α subunit, three residues (H12, D19 and D201) of the PHP domain were selected and substituted with alanine by site-specific mutagenesis. We have taken advantage of the procedure designed for the overexpression of the αεθ complex to purify the αD201A variant.

HoLaMa is an artificial mini-DNA polymerase designed in my host laboratory by enzyme engineering. The interest of artificial enzymes regards the fact that they help to fully understand the catalytic mechanism of DNA polymerases. By creating this mini-DNA polymerase, we obtained an enzyme suitable for NMR studies, lacking the exonuclease activity since the 3' → 5' exonuclease domain of the parent enzyme, the Klenow polymerase, was entirely removed. Therefore, HoLaMa is an enzyme featuring DNA polymerase activity exclusively. Another important property of HoLaMa is the presence of a pre-determined site for fluorophores binding. In this Thesis, one of the variants of this artificial DNA polymerase (lacking the histidine-tag) was overexpressed, purified and kinetically characterized, and its performances were compared with those of HoLaMa provided with the histidine tag. Both artificial polymerases were also compared with Klenow enzyme.

Finally, the last goal of this PhD project is related to the characterization of *E. coli* DNA polymerases I and III at the single molecule level by Förster Resonance Energy Transfer (FRET). These experiments were carried out at Wageningen University (The Netherlands). Nowadays, FRET is presented as one of the most accurate techniques for the analysis of DNA-protein interactions. In addition, the fact of working at the single-molecule level makes possible an approach to the physiological conditions, because of the sorting of different molecules from a heterogeneous population and the possibility of analyze interactions of and between molecules. We have established a procedure for the study of polymerase binding and DNA polymerization. Double-labeled DNA molecules were immobilized on a glass surface. The changes in transfer of energy from a donor dye to an acceptor chromophore upon protein binding and polymerization were measured using different DNA substrates and DNA polymerases. For smFRET experiments two important techniques were combined: the Alternating Laser Excitation (ALEX) and the Total Internal Resonance Fluorescence (TIRF) microscopy. Results obtained by this biophysical analysis might be compared with those obtained by biochemical enzymatic assays.

CHAPTER 2 - MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Bacterial strains

In the present work, only one bacterial strain has been used and transformed:

- *E. coli* TOP10 (F-, *mcrA, Δ(mrr-hsdMRS-mcrBC), ΔlacX74, φ 80lacZDM15, endA1, recA1, araD139, Δ(araA-leu)7697, galU, galK, rpsL*(Str^R), *nupG*) (Invitrogen, Carlsbad, USA) transformed with pBAD-*dnaE*, pBAD-*dnaE*H12A, pBAD-*dnaE*D19A, pBAD-*dnaE*D201A, pGOOD-*dnaQ-hol*E, pGOOD-*dnaX-dnaQ-holE* and/or pBAD-HoLaMa plasmids for protein expression experiments.

2.1.2 Plasmids and genes

The pBADHisB plasmid (Invitrogen) (Fig. 2.1.A) is an ampicillin resistance pBR322-derived expression vector from the ColE1 family, where gene expression takes part in the polylinker region and is controlled by the *araBAD* promoter. In the presence of L-arabinose, expression from pBAD is turned on while the absence of L-arabinose produces very low levels of transcription (Lee *et al.,* 1987). L-arabinose acts as transcription activator and allows optimizing protein expression levels.

pGOOD plasmid (Fig. 2.1.B) was constructed using pTRCHisB (Invitrogen) and pACYC184 (NEB) plasmids (Conte, 2012). The first one is a pBR322-derived expression vector with IPTG as induction agent; the second one contains the p15A origin of replication.

The *dnaE*, *dnaX*, *dnaQ* and *holE* synthetic genes (Entelechon) were cloned into the pBAD (*polA* and *dnaE*) and pGOOD (*dnaX*, *dnaQ* and *holE*) vectors, taking advantage of the NcoI and PstI restriction sites. The pBAD-*dnaE*D201A, pBAD-*dnaE*H12A and pBAD-*dnaE*D19A constructs, which were used to express mutant variants of α subunit, were produced by Entelechon (BadAbbach, Germany) by site-specific mutagenesis. The sequence encoding for HoLaMa was cloned into the pBAD vector using NcoI and PstI sites.
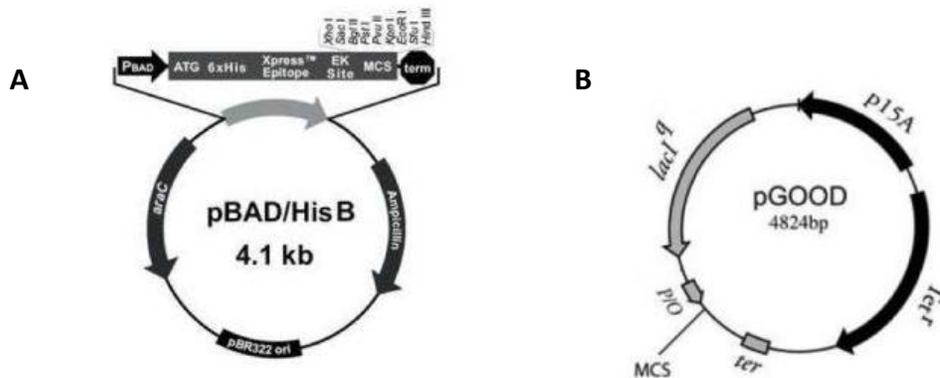
Figure 2.1: pBAD (**A**) and pGOOD (**B**) plasmids for protein co-expression (Conte, 2012).

2.1.3 Media

Luria-Bertani (LB) medium was used for performing bacterial cultures. The composition of this medium was:

- Tryptone 10 g/L
- Yeast extract 5 g/L
- NaCl 10 g/L

Solid medium dishes were prepared by adding agar (15 g/L) to the liquid medium before sterilization, performed at 120° C for 20 minutes. For the specific selection of resistant strains, the following final concentrations of antibiotics were used:

- Ampicillin 100 µg/ml
- Tetracycline 15 µg/ml
- Rifampicin 500 µg/ml

2.1.4 Chromatographic columns

For protein purification, the following chromatographic columns (GE Healthcare) were used:

- Anion exchange Q-Sepharose FF column (1.6 X 25 cm), with a total volume of 50 ml.
- Cibacron Blue 3G affinity column (1.6 X 15 cm), with a total volume equal to 15 ml.
- HiTrap Heparin HP affinity column (1.6 X 2.5 cm), with a total volume of 5 ml.
- Anion exchange Hitrap Q column (7 X 25 mm), with a total volume equal to 5 ml.
- Sephacryl S-300 gel filtration column (1.6 X 70 cm), with a total volume of 140 ml, loaded by using a 2 ml loop injector with a constant flow rate of 0.6 ml/min. The column was calibrated using "HMW and LMW Gel Filtration Calibration Kits"(GE Healthcare).
- HiTrap Desalting column (1.6 x 2.5 cm), with a total volume of 5 ml.

All chromatographies, except the desalting, were carried out using a GradiFrac chromatograph (Pharmacia Biotech).

2.1.5 DNA sequences

DNA constructs for steady-state assays were obtained from Entelechon (Bad Abbach, Germany). As it is shown in paragraph 2.2.19.1, a 15mer primer was annealed with a 40mer template, and only in the case of the study of the polymerase activity of HoLaMa under non-processive conditions, the 15mer primer was annealed to a 16mer template. Single-turnover experiments were carried out using single and double-primed DNAs (section 2.2.21), created by annealing complementary sequences, which were obtained from Eurofins (Germany).

DNA sequences for smFRET experiments were obtained from Iba GmbH (Göttingen, Germany). Sequences featured the corresponding modifications making possible the labeling with donor and/or acceptor dyes. DNA sequences are shown in paragraph 2.2.23.

2.2 METHODS

I – DNA manipulation

2.2.1 Plasmid construction

In order to perform the α subunit over-expression, the pBAD-*dnaE* plasmid (1160 aa), featuring ampicillin resistance and the pBR322-derived origin of replication, was previously constructed (Bressanin *et al.*, 2009).

With regards to the ε, θ and τ proteins expression, their encoding genes (*dnaQ*, *holE* and *dnaX*, respectively) were cloned into the pGOOD vector. The DNA fragment coding for ε243 was excised from pBAD-*dnaQ* (Bressanin *et al.*, 2009) and inserted into the tetracycline resistant pGOOD (NcoI and PstI sites), yielding pGOOD-*dnaQ* (Conte *et al.*, 2011). The *dnaX* and *holE* genes were synthesized (Entelechon GmbH, Regensburg, Germany) and the *holE* gene was cloned into the pGOOD-*dnaQ* plasmid using PstI-KpnI sites, yielding the pGOOD-*dnaQ-holE* construction. On the other hand, the *dnaX* gene was first cloned into the pGOOD vector (NcoI-PstI sites).

The complete *dnaQ-holE* sequence was then linearized with AlwNI and amplified by PCR from the pGOOD-*dnaQ-holE* construct using two specific primers to insert a Shine-Dalgarno (SD) sequence (AGGAGG), which will be located between *dnaX* and *dnaQ-holE* genes. After digestion (KpnI-EcoRI) of both PCR product and pGOOD-*dnaX* plasmid, the final pGOOD construction contained the *dnaX*-SD-*dnaQ-holE* sequence.

The nucleotide sequence of HoLaMa gene reflects that of the *polA* gene, except some site-specific mutations (Martina, 2014). The artificial gene was synthesized and then cloned into the pBAD-HisB plasmid (Invitrogen) by Entelechon-GmbH (Bad Abbach, Germany). The NcoI and PstI restriction sites were used for cloning the HoLaMa sequence.

The amino acid sequence of HoLaMa is the following:

Met G E R N V E E N I E R P L E R V L S R I E R N G V K I D P K V L H N H S E E L T K R L A E L E K K
A H E I A G E E F N L S S T K Q L Q T I L F E K Q G I K P L K K T P G G A P S T S E E V L E E L A L D Y
P L P K V I L E Y R G L A K L K S T Y T D K L P D M I N P K T G R V H T S Y H Q A V T A T G R L S S T
D P N L Q N I P V R N E E G R R I R Q A F I A P E D Y V I V S A D Y S Q I E L R I M A H L S R D K G L L
T A F A E G K D I H R A T A A E V F G C P L E T V T S E Q R R S A K A I N F G L I Y G M S A F G L A R
Q L N I P R K E A Q K Y M D L Y F E R Y P G V L E Y M E R T R A Q A K E Q G Y V E T L D G R R L Y L
P D I K S S N G A R R A A A E R A A I N A P M Q G T A A D I I K R A M I A V D A W L Q A E Q P R V
R M I M Q V H D E L V F E V H K D D V D A V A K Q I H Q L M E N S T R L D V P L L V E V G S G E
N W D Q A H Stop

2.2.2 Preparation of competent cells for electroporation

Electrocompetent cells were prepared by diluting 1:100 an over-night culture in 25 ml of fresh LB medium, with the appropriate addition of antibiotics. Cells were grown with constant shaking (180 rpm) at 37° C, until reaching a density of population equivalent to an absorbance value between 0.5 and 0.7. Cells were then maintained on ice for 30 minutes and centrifuged at 4,500 rpm for 20 minutes at 4° C in an Allegra X-30R Centrifuge (Bechman Coulter, Pasadena, CA, USA), removing the supernatant.

The pellet was subjected to several cycles of washing and centrifugation under the same conditions with a 10% glycerol solution, cooled in ice, suspending the sample in an initial volume equal to the volume of the culture (25 ml) and subsequently reduced to 12, 5 and 2 ml. The pellet obtained at the end of the last washing step was resuspended in a volume of 1 ml 10% glycerol solution, divided into aliquots of 40 µl and stored at -20° C.

2.2.3 Transformation by electroporation and cell stocks preparation

The aliquots of competent cells were thawed and incubated on ice for 5 minutes after the addition of 5 ng of DNA, approximately. Cells suspensions were then transferred into a special cuvette (0.1 cm, Bio-Rad) to be subjected to a potential pulse (1.8 kV, 200 ohms) by a GenePulsar II electroporator (Bio-Rad). Cells were then quickly resuspended in 1 ml of SOC medium and, maintained in ice, transferred to a 1.5 ml Eppendorf tube to be incubated at 37° C and 180 rpm for one hour. The composition of SOC medium was:

- Tryptone 10 g/L
- Yeast extract 5 g/L
- NaCl 10 g/L
- $MgCl_2$ 10 mM
- KCl 2.5 mM
- Glucose 2 g/L

After incubation, 100 µl of cells were transferred to a solid LB medium plate, while the remaining was centrifuged 5 minutes at 10,000 rpm in a Microcentrifuge 5415R (Eppendorf AG, Germany), resuspended in 100 µl of supernatant and transferred to a second solid LB medium plate.

Bacterial stocks were prepared from transformants by streaking on a selective solid medium dish; individual selected colonies were then inoculated in 1 ml of selective medium and grown for 8 hours under shaking (180 rpm) at 37° C. At the end of the growth, the culture was diluted 1:250 in 25 ml of fresh selective medium and maintained over-night under shaking at 37° C. Afterwards, cells were collected by centrifugation (4,500 rpm for 20 minutes) and resuspended in selective medium supplemented with glycerol (15 % v/v) and stored at -20° C.

2.2.4 Extraction of plasmid DNA

The quantitative extraction of DNA plasmids was performed by the Plasmid Midi kit (Qiagen), using the protocol provided by the manufacturer. Once the DNA was resuspended in sterile water, the quantification and assessment of the purity were obtained by determination of DNA absorbance at 230, 260, 280 and 320 nm, using the GeneQuant Pro spectrophotometer (Amersham Biosciences), and by electrophoresis in agarose gels.

2.2.5 Agarose gel electrophoresis

The electrophoresis gel was prepared at a concentration equal to 0.7% (weight/volume) of Molecular Certified Agarose (Bio-Rad) dissolved in TAE buffer (40mM Tris-acetate, 1 mM EDTA). Before loading DNA samples, the 6X DNA Loading Dye (Fermentas) was added, and the race was conducted using TAE buffer in a Bio-Rad electrophoresis chamber, maintaining a constant voltage of 90 V for about 45-50 minutes. At the end of the run, the detection of the bands was carried out by incubating the gel for two hours under shaking in an aqueous solution of 1 g/L ethidium bromide, and analyzing the results with a Molecular Imager Gel Doc (Bio-Rad).

II – Protein manipulation

2.2.6 Comparison of expression and solubility levels between the wild type DNA Pol III proteins and their mutant variants

To verify the expression and solubility of the proteins of interest, induced and non-induced cultures were carried out in parallel. The protein profile of each culture was compared by SDS-PAGE.

Colonies of *E. coli* transformants were grown at 30° C for 9 hours in 1 ml of LB supplemented with appropriate antibiotics. Cultures were then diluted (1:500) into 100 ml of fresh medium and grown overnight at 30° C. The day afterwards, cultures were divided in two equal volumes and a half of them were induced with 1 mM arabinose (pBAD constructions) and 1 mM IPTG (pGOOD constructions) for 2.5 hours at 30° C under constant stirring, letting the other half as non-induced controls. Once the period of induction was finished, all cultures were harvested by centrifugation at 10,000 rpm for 10 min, resuspended in 40 ml of buffer 50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA, 5 mM DTT, 1 mM PMSF and subjected to 3 cycles of sonication with a Misonix Sonicator 3000 (QSonica, Newton, CT, USA). The resulting extract was centrifuged at 10,000 rpm at 4° C for 10 minutes.

For *E. coli* TOP10/pBAD-*dnaE*D201A, various concentrations of detergents and solubilizing agents were added after sonication, shaking the sample (180 rpm) at room temperature for 30 minutes and then centrifuging for 10 minutes at 10,000 rpm at 20° C. The tested agents included:

- 1% v/v Triton X-100

- 0.1% w/v SDS

- 1 % v/v Tween-20

- 1M NaCl

- 40 U DNAse

The soluble fraction of crude protein extracts was appropriately treated as described in paragraph 2.2.14 and loaded on gel electrophoresis (12.5% SDS-PAGE).

2.2.7 Confront of HoLaMa overexpression conditions

A single colony of *E. coli* transformant was cultured at 37° C for 9 hours in 1 ml of LB, ampicillin supplemented. Cultures when then diluted (1:250) in fresh LB medium and grown overnight under constant shaking (180 rpm) at 37° C. The next day, the inoculum was diluted (1:250 dilution) in LB-Amp medium and cultures were incubated over-day at 30° C. Two induction protocols were tested, by adding 1mM arabinose for 6 or 15 h at 30° C, and they were compared by SDS-PAGE.

2.2.8 DNA Pol III overexpression and extraction

One single colony of each bacterial strain was grown at 37° C for nine hours in 1 ml of LB supplemented with suitable antibiotics. Cultures were diluted (1:500) in fresh medium and grown overnight at 30° C. Afterwards, protein overexpression was induced with 1 mM arabinose (for *dnaE* gene) and/or 1 mM IPTG (for genes *dnaQ*, *dnaX* and *holE*) for 2 hours and 30 minutes at 30° C. Then, cells were harvested by centrifugation at 10,000 rpm for 10 minutes at 4° C, resuspended in 50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA, 5 mM DTT, 1 mM PMSF, and subjected to 7-10 cycles of sonication (depending on the amount of harvested cells) at 18 W for 2 minutes each, steps of 15 seconds of pulse and 15 seconds of pause for each cycle. Between two cycles, the

sample was cooled on ice for 5 minutes. At the end of the sonication process, samples were centrifuged at 10,000 rpm for 10 min and the soluble fraction was collected and filtered.

2.2.9 α wild type subunit, αH12A, αD19A and core purification

The purification of the α subunit, two of its mutant variants (αH12A and αD19A) and the core was obtained by standard chromatographic techniques.

The soluble fraction obtained after the extraction process was loaded onto the anion exchange **Q-Sepharose column** (1.6 x 25 cm), equilibrated with buffer 50 mM Tris–HCl pH 8, 50 mM NaCl, 1 mM EDTA, 2.5 mM DTT. After loading the sample (1 ml/min flow rate), the column was washed with 5 column volumes of equilibration buffer. The elution was performed with a gradient of NaCl up to 600 mM (8 column volumes in buffer 50 mM Tris–HCl pH 8, 2 M NaCl, 1 mM EDTA, 2.5 mM DTT), at a flow rate of 2 ml/min. Collected fractions (6 ml each one) were diluted 1:2 using the buffer 50 mM Tris–HCl pH 8, 1 mM EDTA, 2.5mM DTT, 40 % glycerol, to halve the concentration of NaCl and achieve a final concentration of 20% glycerol.

Fractions containing α subunit, its mutant variants or the αεθ were then pooled and concentrated to a final volume around of about 40 ml, using an Amicon Stirred Ultrafiltration Cell fitted with a YM100 membrane (Millipore Corporation, Billerica, MA, USA), proceeding with serial dilutions with a 50 mM Tris–HCl pH 8, 50 mM NaCl, 1 mM EDTA, 2.5mM DTT, 20 % glycerol buffer, in order to bring the final concentration of NaCl to 50 mM, approximately. Sample was centrifuged at 10,000 rpm for 10 minutes at 4° C and the obtained soluble fraction was supplemented with 5 mM $MgCl_2$ and loaded onto an affinity **Cibacron blue column** (1.6 x 15 cm) at a flow rate of 0.5 ml/min. This column was previously conditioned with 50 mM Tris-HCl pH 8, 50 mM NaCl, 2.5 mM DTT, 20 % glycerol, 5 mM $MgCl_2$ buffer. After loading the sample, the column was washed with 5 column volumes of equilibration buffer and, then, the elution was performed in two steps (150 mM NaCl and 1 M NaCl) with 50 mM Tris–HCl pH 8, 2 M NaCl, 2.5 mM DTT, 5 mM $MgCl_2$, 20 % glycerol buffer, at a flow rate of 2 ml/min. The second peak of elution was separated in three pools called "head" (3 ml), "core" (20 ml) and "tail" (20 ml). To these samples, 50 mM EDTA was added as $MgCl_2$ chelating agent.

The "core" and "tail" samples were pooled, diluted with buffer 50 mM Tris–HCl pH 8, 1 mM EDTA, 2.5 mM DTT, 20 % glycerol and concentrated at a final concentration of NaCl equal to 50 mM, by ultrafiltration with Amicon Stirred Ultrafiltration Cells, fitted with YM100 membranes. The sample was then loaded onto a **HiTrap Heparin HP column** (0.7 x 2.5 cm), equilibrated with 50 mM Tris–HCl pH 8, 50 mM NaCl, 1 mM EDTA, 2.5 mM DTT, 20 % glycerol buffer. After protein loading (0.5 ml/min flow rate), the column was washed with 5 column volumes of equilibration buffer and, then, a NaCl gradient up to 300 mM NaCl (10 column volumes) was applied at a flow rate of 1 ml/min. Fractions of 0.9 ml were collected, and those containing the proteins of interest were pooled and concentrated with Amicon Stirred Cells to reduce the concentration of NaCl.

Finally, the α and αεθ samples were loaded onto an anion exchange **Hitrap Q column** (7 x 25 mm) at a flow rate of 0.5 ml/min. Equilibration and elution buffers were the same used for the Hitrap Heparine affinity column. The final concentration of NaCl at the end of the gradient was set equal to 600 mM, reached in 50 ml. Fractions of 0.9 ml were collected and those featuring a high concentration of the protein of interest were pooled and concentrated.

2.2.10 Complexes purification

The overexpression, extraction and purification protocols followed in the case of the pBAD-*dnaE*-pGOOD-*dnaX-dnaQ-holE* complex (αεθ) and its mutant variant, pBAD-*dnaE*D201A-pGOOD-*dnaX-dnaQ-hol*E (αD201Aτεθ), were the same as described above for the catalytic core.

Initially, a first gel filtration of the crude extract pBAD-*dnaE*-pGOOD-*dnaX-dnaQ-holE* complex, obtained by sonication of 3 L of culture, was performed so as to elucidate the type of protein sub-assembly. For this purpose, the pellet yielded from three liters of cell culture, harvested at 10,000 rpm for 10 minutes at 4° C, was diluted in 25 ml of buffer 50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA and 1 mM PMSF and sonicated 3 times, as described in paragraph 2.2.8. After centrifugation at 10,000 rpm for 10 minutes at 4° C and filtration of the soluble extract, 1 ml was loaded onto a Sephacryl S-300 column (GE Healthcare, 1.6 x 62 cm), equilibrated with 50 mM Tris–HCl pH 8, 150 mM NaCl, 1 mM EDTA, 2.5 mM DTT, 20% glycerol, using a loop of 1 ml. The flow rate was 0.6 ml/min, and fractions of 0.9 ml were collected. All fractions were then subjected to enzymatic assays (2.2.20) and protein quantification was obtained by the Bradford assay (2.2.13).

Besides, a complete purification protocol was developed for 12 liters of cell culture, both for pBAD-*dnaE*-pGOOD-*dnaX-dnaQ-holE* and pBAD-*dnaE*D201A-pGOOD-*dnaX-dnaQ-holE* complexes, following the same procedure used in the case of α subunit or the catalytic core. The exception was the Cibacron blue affinity column, where the first washing step with 150 mM NaCl was not applied. Moreover, after carrying out the purification step with the HiTrap Heparin affinity column, a gel filtration process was performed using the Sephacryl S-300 column so as to determine, once again, the identity and the type of protein sub-assemblies present for both purified complexes.

Fractions eluted from the HiTrap Heparin column were concentrated with an Amicon Stirred Ultrafiltration Cells, and 2 ml were loaded onto a Sephacryl S-300 column (GE Healthcare, 1.6 x 62 cm), equilibrated with 50 mM Tris–HCl pH 8, 150 mM NaCl, 1 mM EDTA, 2.5 mM DTT, 20% glycerol, at 0.6 ml/min flow rate. The gel filtration column was previously calibrated with HMW and LMW Gel Filtration Calibration kits (GE Healthcare). The volume of the fractions was equal to 0.9 ml. Fractions of interest were pooled and concentrated without performing a Hitrap Q anion exchange column.

2.2.11 HoLaMa overexpression and extraction

One individual colony was resuspended in 1 ml of LB-Amp medium and grown under shaking at 180 rpm and 37° C for nine hours. This pre-culture was diluted 1:250 in LB medium, ampicillin supplemented, and grown overnight under constant stirring (180 rpm) at 37° C. The following day, the inoculum was diluted (1:250 dilution) in LB-Amp medium and cultures were incubated over-day at 30° C. The induction was triggered by 1mM arabinose and held for 15 h at 30° C. Cells were collected and sonicated following the same protocol used for α subunit, omitting DTT from the resuspensión buffer.

2.2.12 HoLaMa purification

The purification procedure of HoLaMa was similar to that of α subunit or the catalytic core. The protocol included an anion exchange Q-Sepharose FF, a Cibacron Blue affinity column and, finally, an affinity HiTrap Heparin HP chromatography. Only three major points differed from the protocol used for purification of α subunit:

- No DTT was present in any buffer during the purification protocol due to the fact that the presence of just one cysteine makes unlikely aggregation by disulfide bonds.

- Only one eluting step with 1 M NaCl was used during the Cibacron blue chromatography, in order to avoid the loss of protein during the used washing step with 150 mM NaCl.

- The concentration of pooled fractions between two steps of purification was performed by serial dilutions and using an Amicon Stirred Ultrafiltration Cells fitted with a YM30 membrane (Millipore Corporation, Billerica, MA, USA), instead of a YM100 membrane, according to the molecular mass of HoLaMa (46 kDa).

2.2.13 Protein quantification

The measurement of protein concentration was performed according to the method of Bradford (Bradford, 1976), assessing the increase of absorbance of the Coomassie Blue G-250 dye at 595 nm, as a result of protein binding. Calibration was performed using suitable dilutions of a standard solution of bovine serum albumin. Each determination was performed in duplicate using a Microplate reader 550 (BioRad)

2.2.14 SDS-PAGE analysis

The analysis of protein samples was carried out with 10% (HoLaMa) or 12.5% (DNA Pol III) acrylamide gels prepared with Bio-Rad (Hercules, CA, USA) reagents, according to Table 3. Protein samples were diluted (1:5) in a 5X denaturing loading buffer, composed of 500 mM Tris-HCL pH 6.8, 2.7 M glycerol, 10% SDS, 0.5% bromophenol blue and 5% β-mercaptoethanol, and then boiled for 5 minutes, immediately before loading. The electrophoresis was carried out in a Mini-Protean cell (Bio-Rad) for 90 minutes, approximately, at constant voltage (120 V) and using as running buffer 25 mM Tris pH 8.3, 200 mM glycine, 5 mM SDS. Protein bands were visualized by gel incubation under constant shaking in Coomassie Brilliant Blue R-250 for 40 minutes, followed by two stages of destaining, 40 min each one, in a 30% ethanol-10% acetic acid solution.

Reagent	Running Gel	Stacking gel
Acrylamide	10% or 12.5%	4%
Tris-HCl pH 8.8	0.25 M	-
Tris-HCl pH 6.8	-	0.25 M
SDS	0.2%	0.2%
TEMED	0.1%	0.1%
APS	0.25%	0.25%

Table 3: List of reagents for preparing 10% or 12.5% SDS-PAGE.

2.2.15 Protein cross-linking

The trimeric purified ατεθ complex was subjected to cross-linking with bis(sulfosuccinimidyl)suberate (BS3) (ThermoScientific, Waltham, MA, USA) with the purpose of identifying interactions between subunits. Previously, a desalting step with a HiTrap Desalting column (GE Healthcare), filled with a Sephadex G-25 Superfine matrix (GE Healthcare), was performed so as to remove Tris-buffer from the sample, as this amine could interact with the cross-linker. Protein sample (1.5 ml) was injected after equilibrating the column (1.6 x 2.5 cm) with 5 volumes of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). Column was then washed with PBS and, after the elution of the first 1.5 ml, the following 2 ml of the eluted sample were collected and concentrated to 350 nM with an Amicon Microcel 100K (Merck Millipore, Darmstadt, Germany) by several cycles of centrifugation, 4 minutes each, at 3,700 g and room temperature.

The trimeric replicase (350 nM) was incubated with 5, 50, 500 or 1000 µM BS3 in PBS for 1 hour at room temperature. Reactions were stopped with 100 mM Tris pH 7.9. Untreated ατεθ complex was used as control. Then, samples were resuspended (1:5) in a denaturing loading buffer, boiled for 5 minutes and loaded onto a 10% Polyacrylamide gel, as described in paragraph 2.2.14.

2.2.16 Mass spectrometry of the trimeric replicase

Mass Spectrometry was performed as previously described (Shevchenko *et al.,* 2006; Conte *et al.,* 2012) by Electrospray Ionization mass spectrometry (ESI-MS) using a Q-Tof Premier Instrument coupled with a nanoAcquity UPLC (Waters, Milford, USA). Samples were protein bands from SDS-PAGE gels subjected to digestion with trypsin. Data were introduced in the Mascot software (Matrix Science) with the National Center for Biotechnology Information (NCBInr) database information. Criteria used in mass spectrometry were: mass accuracy window for parent ion, 50 ppm; mass accuracy window for fragment ions, 200 millimass units; fixed modification, carbamidomethylation of cysteines; variable modifications, oxidation of methionine.

2.2.17 Bacterial growth kinetics

Growth kinetics of the TOP10/pBAD-α, TOP10/pBAD-αH12A and TOP10/pBAD-αD201A strains were spectrophotometrically determined by following the change in absorbance at 600 nm of six different cell cultures (both induced and non-induced strains). The procedure started with the inoculum of a streaked transformant colony in 1 ml of LB-ampicillin. The pre-cultures, grown over-day at 37° C under constant shaking (180 rpm), were diluted (1:500) in LB-ampicillin and maintained at 37° C over-night under constant shaking. The cultures were then transferred to different flasks, diluted and a half of them was induced with 1 mM arabinose. Transmittance determination was carried out using a turbidimeter Biolog 21901 (Hayward, CA, USA).

During bacterial growth, the wild type and mutant host cells were subjected to phenotypic examination, with the aid of a fluorescence microscope and Hoechst dye, suitable for fluorescence DNA analysis, to determine differences between strains. Furthermore, induced and non-induced cultures were plated on LB dishes, antibiotic free, for growth title determination at fixed times: 0, 4, 4:40, 5:20, 6, 8 and 10:30 hours after induction. 100 µl were plated on any LB dish.

2.2.18 Mutator phenotypes assays

One colony of *E. coli* TOP10/pBAD-*dnaE* and two of its mutant variants (TOP10/pBAD-*dnaE*D201A, TOP10/pBAD-*dnaE*H12A) were grown at 37° C for ten hours in 1 ml of LB, ampicillin supplemented. 100 µl from every pre-culture were transferred into 5 ml of LB-ampicillin and maintained at 37° C

over-night under constant shaking. The following day, a new 1:500 dilution was performed in 20 ml of 1 mM arabinose-induced LB-ampicillin cultures, and preserved over-day at 37° C under constant shaking. Non-induced cultures were used as controls. At 0, 4, 6 and 10 hours after the induction, 100 µl of any sample were plated on LB or LB-Rifampicin (500 µg/ml) plates. The LB plates were used to determine population density whereas rifampicin plates allow measuring mutation frequency. Plates were incubated for 24 hours at 37° C, and the appearance of single colonies was examined. Both the title and the frequency of mutation were calculated as the average number of single colonies in three serial dilutions from independent cultures.

III - Enzymatic activity assays

2.2.19 Standard steady-state activity assays

For enzymatic activity determinations, a Perkin -Elmer λ19 spectrophotometer with a 1 cm quartz cuvette (final volume of 1 ml) was used. Enzymes used in the coupled assays were purchased from Sigma-Aldrich.

2.2.19.1 DNA polymerase activity assay

In order to determine the polymerase activity of α subunit, a coupled-enzyme method developed in my host laboratory by Suarez *et al.* (2012) was used. The assay takes advantage of the inorganic phosphate produced by the cleavage of pyrophosphate, arising from the polymerase activity of α subunit, by the inorganic pyrophosphatase (PPase). This inorganic phosphate allows the development of a series of coupled enzymatic reactions, indicated in Fig. 2.2, which lead to obtain a molecule of uric acid as a final product. In the presence of inorganic phosphate, the purine nucleoside phosphorylase (PNPase) catalyzes inosine phosphorolysis, producing hypoxanthine that is oxidized to uric acid by the action of the xanthine oxidase (XOD). The uric acid absorbance can be monitored at 293 nm according to a molar extinction coefficient that does not depend on pH values, in the range of 6 and 9.
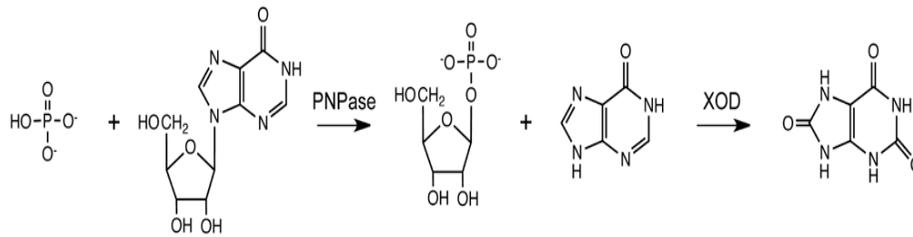
Figure 2.2: Coupled-enzyme reactions that lead to the production of uric acid from the phosphate released after pyrophosphate hydrolysis (Suárez *et al.,* 2012).

The polymerase reaction substrate is a double-stranded DNA, which allows the extension of 25 nucleotides, from the annealing of a 40mer DNA template and a 15mer DNA primer, both produced by Entelechon (Bad Abbach, Germany).

```
TEMPLATE:  3'-CGCGCGCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-5'
PRIMER:    5'-GCGCGCGCTTTTTTT-3'
```

For HoLaMa polymerase assay, another DNA substrate was also used: a double-stranded DNA with only one extending nucleotide, achieved by annealing a 16mer template to a 15mer primer (Entelechon, Bad Abbach, Germany). This DNA allowed us to evaluate kinetic efficiency in the absence of processivity. In this case, Klenow enzyme was used as a control.

```
TEMPLATE:  3'-CGCGCGCGAAAAAAAA-5'
PRIMER:    5'-GCGCGCGCTTTTTTT-3'
```

Reactions were started by the addition of 100 μM dTTP and were monitored at 293 nm. The concentration of the wild type α subunit and the D19A mutant variant was equal to 17 nM, whereas the H12A mutant variant was used at 13 nM; the ατεθ complex concentration was 5 nM, whereas the αD201Aτεθ complex was equivalent to 3.2 nM. The final concentration of the components of the polymerase activity reaction mixture was:

- 100 mM Tris pH 8
- 5 mM $MgCl_2$

- 1 µM dsDNA
- 0.25 mM inosine
- 40 mU/ml PPase
- 50 mU/ml PNPase
- 500 mU/ml XOD
- 100 µM dTTP

Polymerase activity was also evaluated in the absence of an exogenous pyrophosphatase so as to determine the relationship between both polymerase and the hypothesized intrinsic pyrophosphatase activity of the PHP domain (2.2.19.3).

2.2.19.2 Pyrophosphatase activity assay

To detect the pyrophosphatase activity, pyrophosphate was used as substrate (1 mM final concentration) and the phosphate product was quantified by absorbance measurements at 293 nm. Obviously, PPase was absent to test the hypothetical pyrophosphatase activity of the PHP domain. Enzyme concentration was equal to 17 nM in the case of the wild type α subunit and the D19A mutant variant, whereas the H12A mutant variant was used at 13 nM; both αεθ and αD201Aτεθ complexes were used at a concentration equal to 4.5 nM. The final concentration of the reagents of the pyrophosphatase assay reaction mixture was:

- 100 mM Tris pH 8
- 10 mM $MgCl_2$
- 250 µM $MnCl_2$
- 0.25 mM inosine
- 50 mU/ml PNPase
- 500 mU/ml XOD
- 1 mM sodium pyrophosphate.

2.2.19.3 Associated polymerase and pyrophosphatase activities

The intrinsic pyrophosphatase activity directly co-related to the degradation of each molecule of pyrophosphate released after the incorporation of each nucleotide during the polymerization process was evaluated ensuring the conditions that enable both activities to be developed. Both Mg^{+2} and Mn^{+2} were present and the assay was carried out in the presence and absence of exogenous pyrophosphatase. No pyrophosphate was added to the reaction mix just to determine the concentration of pyrophosphate that is hydrolyzed as a consequence of the polymerase reaction. Enzyme concentration was the same than in the case of the polymerase assay. The final concentrations of the reaction mixture were:

- 100 mM Tris pH 8
- 5 mM $MgCl_2$
- 10 mM $MnCl_2$
- 0.25 mM inosine
- 1 µM dsDNA
- 40 mU/ml PPase
- 50 mU/ml PNPase
- 500 mU/ml XOD
- 100 µM dTTP

2.2.19.4 Exonuclease activity assay

Exonuclease activity was assayed according to Hamdan *et al.* (2002b), using the 5'-p-nitrophenyl ester of thymidine monophosphate (pNP-TMP) as substrate (Fig. 2.3). The exonuclease activity of 5.2 nM α subunit, 5.2 nM αεθ and 1.8 nM ατεθ complex was monitored at 420 nm, maximum absorption value of the p-nitrophenolate released from the p-nitrophenyl ester of the pNP-TMP. The assay mixture contained:

- 100 mM Tris pH 8
- 1 mM $MnCl_2$
- 3.3 mM pNP-TMP
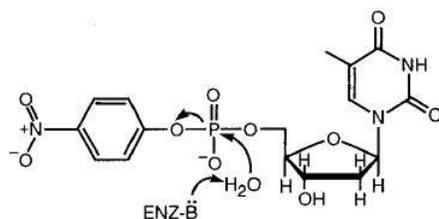
Figure 2.3: Mechanism of hydrolysis of the pNP-TMP substrate for 3' → 5' exonuclease activity assay (Hamdan *et al.,* 2002b)

2.2.20 Steady-state activity assay in microplate

As it was previously described in paragraph 2.2.10, an initial gel filtration chromatography of the crude extract of the pBAD-*dnaE*-pGOOD-*dnaX-dnaQ-holE* complex was developed. Specific assays for each subunit were performed, with the purpose of establishing a relationship between the elution peaks observed on the chromatogram and the types of protein sub-assemblies. These assays were carried out using a Microplate reader 550 (BioRad). Each reaction was developed for all eluted fractions in a final volume of 200 µl. The coupled-enzyme assay, above described, was again used but using the 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-phenyl-2*H*-tetrazolium (INT)-formazan as final electron acceptor instead of oxygen. Absorbance measurement was performed at 460 nm instead of at 293 nm (uric acid).

2.2.20.1 ATPase activity assay

ATPase activity of τ subunit was assayed according to Tsuchihashi and Kornberg (1990). The inorganic phosphate released from ATP hydrolysis, together with the inosine and the same coupled enzymatic reactions of PNPase and XOD, above-described, lead to hypoxanthine oxidation and the consequent reduction of the INT. Reactions were started by the addition of 0.4 mM ATP and monitored at 460 nm. Reaction mixtures contained the following reagents, added to an aliquot of each eluted fraction:

- 100 mM Tris-HCl pH 8
- 5 mM MgCl2

68

- 0.25 mM inosine
- 1 µM ssDNA
- 50 mU/ml PNPase
- 500 mU/ml XOD
- 1 mM INT
- 400 µM ATP

2.2.20.2 Polymerase activity assay

Polymerase activity was determined using the enzyme-coupled assay, previously described. The reaction mixture was prepared by adding the following reagents to an aliquot of each eluted fraction:

- 100 mM Tris-HCl pH 8
- 5 mM $MgCl_2$
- 1 µM dsDNA
- 0.25 mM inosine
- 10 mU/ml PPase
- 50 mU/ml PNPase
- 500 mU/ml XOD
- 1 mM INT
- 100 µM dTTPs

2.2.20.3 Exonuclease activity assay

The reaction mixture was mixed with a variable aliquot of each fraction, collected after chromatography. The assay mixture contained the following reagents:

- 100 mM Tris-HCl pH 8
- 1 mM $MgCl_2$
- 3 mM pNP-TMP

2.2.21 Inter-primer distance effect and single turnover activity assay

Three double-primed DNAs (dpDNA) were used as substrate (100mer, 75mer and 50mer). In all cases, dTTP and dATP were required for the extension close to the 3' or 5' end of the template, respectively. Moreover, these DNAs featured 65, 40 and 15 nucleotides of inter-primer distance, in order to determine the effect this parameter on the simultaneous replication. On the other hand, a single-primed 25mer DNA (spDNA), whose elongation depended on dGTP, was also selected. Hence, dpDNA served as a model of lagging strand while the spDNA mimicked the leading strand. DNA substrates used for single turnover assays and inter-primer evaluation were:

```
25mer: 5' CCCCCCCCCCCCCCCGGGGGGGGGG 3'
        3'                  CCCCCCCCCC 5'


50mer: 5'TTTTTTTTTTTTTTTGGGGGGGGGGGAAAAAAAAAAAAAAAACCCCCCCCCC 3'
        3'              CCCCCCCCCC              GGGGGGGGGG 5'


75mer: 5'TTTTTTTTTTTTTTTGGGGGGGGGGGAAAAAAAAAA…AAAAAAAAAACCCCCCCCCC 3'
        3'              CCCCCCCCCC                      GGGGGGGGGG 5'


100mer: 5'TTTTTTTTTTTTTTTGGGGGGGGGGGAAAAAAAAAAAAAAAAA…AAAAAAAAAACCCCCCCCCC 3'
         3'              CCCCCCCCCC                          GGGGGGGGGG 5'
```

The effect of the inter-primer distance on DNA elongation was evaluated under steady state conditions: 92 nM α subunit, 100 μM dATTP, dGTP or dTTP and 1 μM 25mer, 50mer, 75mer or 100mer DNA substrate. Polymerase activity was assayed using a stirred QS Hellma quartz cuvette (1mm light path, 350 μl total volume) and a Jasco V-550 spectrophotometer. Single turnover assays were performed both for α subunit and the trimeric αεθ complex. Assay and detection conditions were the same as for the polymerase standard steady-state determinations (described in paragraph 2.2.19.1), except for substrate and enzyme concentrations, both at 0.2 μM. In addition, to ensure minimization of the lag time, PPase, PNPase, and XOD were used at 40, 250 and 500 mU/ml, respectively.

2.2.22 Pyrophosphatase activity inhibition assay

The inhibition assay of the inorganic PPase (0.45 nM) by the fluoride ion was performed by the following reaction mixture:

- 100 mM Tris pH 8
- 10 mM MgCl$_2$
- 0.25 mM inosine
- 200 mU/ml of PNPase
- 500 mU/ml of XOD
- 5 mM pyrophosphate
- 24 mU/ml of PPase.

Different concentrations of fluoride ion (20, 40, 100, 200, 400, 800 and 1200 µM) were added to the solution mixture. The assay was then repeated replacing the inorganic PPase with the purified wild-type α subunit (30 nM), in order to study and confirm the hypothetical pyrophosphatase activity of the PHP domain. Once again, different concentrations of fluoride ion (20, 50, 200, 800 and 1200 µM) were added to the reaction mixture. Following the effective hydrolysis of pyrophosphate into two molecules of phosphate, the reaction continued with the same coupled enzyme reactions, above-described for previous assays, determining the absorbance of the reaction final product, uric acid, at 293 nm.

IV-smFRET experiments

Single-molecule Förster Resonance Energy Transfer (smFRET) allows the determination of distances between two fluorophores in the nanometre range. FRET is based on a non-radiative energy transfer from a donor fluorophore to an acceptor chromophore (Förster, 1946). Three conditions are essential for FRET to occur:

- Donor emission and acceptor absorption spectra must overlap;
- Donor and acceptor dyes ought to be close to each other (<10 nm);
- The relative orientation of the donor and the acceptor transition dipole moments must allow transfer of energy (Lakowicz, 2006).

Basically, the DNA molecule is labeled at two points and laser beam is focused via an objective on the molecules in order to excite the donor probe. The absorption of the photon emitted from the donor dye in the acceptor dye will depend directly on the distance between both dyes. As the distance between dyes will change as a result of protein binding and/or polymerization, we can determine molecular dynamics. SmFRET experiments have been carried out at the Laboratory of Biophysics of Wageningen University (The Netherlands).

2.2.23 DNA sequences

Single-strand DNA sequences were obtained from Iba GmbH (Göttingen, Germany):

DNA SEQUENCES

003: 5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC$_H$ 3'

004: 5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'

025: 3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA TC**X** CGG ACG AAG CAC C 5'

026: 3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA **X**CT CGG ACG AAG CAC C 5'

027: 3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC **X**CA TCT CGG ACG AAG CAC C 5'

028: 3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT C**X**C TCA TCT CGG ACG AAG CAC C 5'

032: 3' *p*TA GAG AGT AGA GCC TGC TTC GTG G 3'

040: 3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA TCT CGG ACG AAG CAC **X** 5'

Table 4: DNA sequences used for DNA annealing in smFRET. **T** represents the acceptor position; **X** represents the donor dye position; C$_H$ indicates 3'-terminus dideoxi-terminated; *p* indicates 5' phosphate group.

Both DNA 003 and 004 are 30 bases long and contain a C6-Amino-2'-deoxythymidine for labeling with ATTO647N (T), used as acceptor dye. Both DNAs were biotininylated on the 5'-terminus to allow immobilization of the DNA using neutravidin as a linker with biotin. Unlike 004, the 003 sequence bears a 3'-terminus dideoxi-terminated (C_H 3'), which blocks the polymerization process and restricts the use of DNA 003 to binding experiments. The rest of the sequences (025, 026, 027, 028 and 040) were complementary to 003 and 004 strands and featured a C6-Amino-2'-deoxythymidine, at different position depending on the sequence, allowing the labeling with the Cy3B donor dye (X). DNA 032 lacked modified bases for labeling purposes and was complementary to 025, 026, 027, 028 and 040 strands.

2.2.24 DNA labeling

For optimal DNA precipitation, 5 nmoles DNA sequence were mixed with 5 µl of 3 M sodium acetate and 125 µl ice-cold 100% ethanol and stored at **-20°** C overnight. Samples were then centrifuged for 30 minutes at 4° C and 16,000 g and the supernatant was removed. Pellets were washed with 200 µl ice-cold 70% ethanol and centrifuged at 4° C and 16,000 g for 15 minutes. After removing the supernatant, the pellets were dried at 40° C and latter dissolved in 95 µl of 0.1 M sodium borate pH 8.5 before 50 nmoles of Cy3B (donor dye) or ATTO647N (acceptor dye), dissolved in 5 µl DMSO, was added. Samples were wrapped in aluminium foil to protect them from light and placed for 1 hour on a table-top orbital shaker. Another 50 nmoles of dye was added to the DNA and, after mixing the solution properly, the samples were stored in a cold room at 4° C overnight. The next day, 10 µl of 3 M sodium acetate and 125 µl ice-cold 100% ethanol were added to the reaction mix to precipitate the DNA and the samples were stored at -20° C overnight. The following day, samples were centrifuged for 30 minutes at 4° C and 16,000 g and the supernatant was removed. The pellet was rinsed with 200 µl ice-cold 100% ethanol and centrifuged for 15 minutes at 4° C and 16,000 g to remove the supernatant. Once dried, labeled samples were kept frozen.

2.2.25 ssDNA purification using denaturing urea-polyacrylamide gel

This denaturing gel was prepared according to the following Table:

Reagent	Amount / volume
6M Urea	14.4 g
Acrylamide 40%	20 ml
TBE 10x	4 ml
TEMED	27 µl
10% APS	270 µl

Table 5: Reagents for urea-20% polyacrylamide gel for ssDNA purification

Milli-Q water was added up to a volume of 40 ml. Pellets of labeled DNA were resuspended in 10 µl TE buffer (Tris 10 mM pH 8, EDTA 1mM) and 10 µl formamide (J.T. Baker, Center Valley, PA, USA). After heating the samples to 95° C for 5 minutes, 20 µl were loaded on each lane of the Urea – 20% polyacrylamide gel. Empty lanes were filled with 20 µl of 1:1 formamide-xylene cyanol and electrophoresis was conducted in a LSG-400-20 chamber (CBS Scientific, San Diego, CA, USA) at 300 V for around 4 hours, always under temperature control (less than 50° C). 1x TBE (Tris-borate-EDTA) (Sigma-Aldrich, St. Louis, MO, USA) was used as running buffer. An UV lamp (250 nm wavelength) was used to cast a shadow from the DNA onto the UV sensitive TLC plates RP 18F$_{254}$S (Merck, Whitehouse Station, NJ, USA). The shadow represented the labeled DNA. Labeled DNA bands were cut and subjected to the 'crush and soak' method: gel fragment was crushed and soaked in 20 volumes of TE buffer, shacking overnight at room temperature. The next day, gel mixtures were centrifuged in the cold room (4° C) for 10 minutes at 16,000 rpm and the supernatant was taken to precipitate the DNA by adding 0.1 volumes 3M sodium acetate and 2.5 volumes ice cold 100% ethanol. Samples were mixed and stored at -20° C overnight. The next day, samples were centrifuged for 30 minutes at 16,000 rpm and 4° C. Pellets were rinsed with 200 µl ice-cold 70% ethanol. Successively, samples were centrifuged for 15 minutes at 4°C and 16,000 rpm and pellets were dried at 40° C. Then, each pellet was dissolved in 20 µl TE buffer and DNA concentration was determined using Nanodrop 2000c machine (Thermo Scientific, Waltham, MA, USA). The ratio between concentration of DNA and dye was also determined.

2.2.26 DNA annealing

DNA constructions were prepared by annealing complementary sequences for binding and polymerization analysis (Table 6).

DNA SUBSTRATES
004-025   5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'   3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA TC**X** CGG ACG AAG CAC C 5'
004-026   5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'   3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA **X**CT CGG ACG AAG CAC C 5'
004-027   5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'   3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC **X**CA TCT CGG ACG AAG CAC C 5'
004-028   5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'   3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT C**X**C TCA TCT CGG ACG AAG CAC C 5'
004-040   5' CCT CAT TCT TCG TCC CAT **T**AC CAT ACA TCC 3'   3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA TCT CGG ACG AAG CAA **X** 5'

Table 6: DNA substrates used for endpoint assays by smFRET experiments. **T** represents acceptor dye, whereas **X** represents donor dye.

Each annealing was performed as follows: 0.5 μM of 004 DNA sequence, 0.7 μM of 025, 026, 027,028 or 040 DNA sequence and 4 μl of annealing buffer (10 mM Tris pH 8, 1 mM EDTA, 250 mM NaCl) was added and topped up with Milli-Q water to a final volume of 40 μl. Samples were placed in boiling-hot water and left until the water reached room temperature. The annealed DNA samples were finally stored at -20° C.

Real time-polymerization and endpoint assays were performed using primer/template constructions bearing a single-stranded DNA overhang. Besides, protein-DNA binding was also studied by using one nucleotide gapped-DNA binding sensor (003-025-032) for Klenow Fragment and HoLaMa protein. The construction included another DNA sequence (032) complementary to

the 025 sequence and with a phosphate group (**p**) in the 5'-terminus. The sequence of the binding sensor used for the Klenow Fragment and HoLaMa titration was:

003-025-032

```
5' CCT CAT TCT TCG TCC CAT TAC CAT ACA TCC_Hp TA GAG AGT AGA GCC TGC TTC GTG G 3'
3' GGA GTA AGA AGC AGG GTA ATG GTA TGT AGG AAT CTC TCA TCX CGG ACG AAG CAC C 5'
```

2.2.27 Endpoint polymerization assays

The polymerization assay was carried out with the aim of screening for the maximum FRET changes between the initial state (long ssDNA overhang) and the final state (dsDNA). For endpoint assays, 10 nM DNA (DNAs 004-025, 004-026, 004-027, 004-028 or 004-040), 100 μM dNTPs and 40 nM Klenow Fragment (KF) were mixed. Reactions were maintained one hour at room temperature and stopped adding 0.1 M EDTA to each sample. Negative controls were prepared, in which we left out the 100 μM dNTPs and replaced the volume with 2 μl of TE buffer. Mastermix was composed of 11.62 nM DNA and 0.79 mM Trolox. Its final volume (17.2 μl) was reached by adding KF7 buffer (50 mM Tris pH 7.5, MgCl$_2$ 10 mM, BSA 100 μg/ml, Glycerol 5%, DTT 1mM).

DNA substrate	Mastermix	TE buffer	dNTPs (stock: 1 mM)	KF (stock: 1 μM)
004-025 (-)	17.2 μl	2 μl	-	0.8 μl
004-025	17.2 μl	-	2 μl	0.8 μl
004-026 (-)	17.2 μl	2 μl	-	0.8 μl
004-026	17.2 μl	-	2 μl	0.8 μl
004-027 (-)	17.2 μl	2 μl	-	0.8 μl
004-027	17.2 μl	-	2 μl	0.8 μl
004-028 (-)	17.2 μl	2 μl	-	0.8 μl
004-028	17.2 μl	-	2 μl	0.8 μl
004-040 (-)	17.2 μl	2 μl	-	0.8 μl
004-040	17.2 μl	-	2 μl	0.8 μl

Table 7: Endpoint assays mixtures. Assays were carried out for 1 hour at room temperature in eppendorf tube. Negative controls (-) lacked dNTPs. Final concentrations: 10 nM DNA, 100 μM dNTPs and 40 nM KF.

For constructs 004-025 and 004-040, the endpoint polymerization assays were repeated with either 40 nM of Klenow Fragment, HoLaMa, α subunit or αεθ, 10 nM 004-025 DNA substrate and 100 μM four dNTPs (dTTP, dGTP, dCTP and dATP), following the same precedent protocol. Negative controls were performed by substituting 100 μM dNTPs for TE buffer. Reactions were carried out for one hour at room temperature and stopped by adding 0.1 M EDTA. DNA molecules were fixed on modified glass slides, coated with vectabond and PEG, and observed using imaging buffer, as described in the following paragraphs.

2.2.28 Vectabond coating

Slides were burned for one hour at 500° C in a muffle furnace (Vecstar, Chesterfiel, UK). After cooling down to room temperature, the slides were transferred into a glass jar and immersed in 50 mL of acetone (Merck-Millipore, Darmstadt, Germany) for 5 min. 1mL of Vectabond solution (Vector Laboratories, Burlingame, CA, USA) was added into the jar for five minutes. The Vectabond-acetone solution was removed and replaced with fresh acetone. Three more washing steps with 50 mL milliQ water followed. Slides were dried using compressed nitrogen.

2.2.29 PEG coating

Once slides were dried, sticky-slides VI (Ibidi, Munich, Germany) were placed over them. For any channel well, 50 μL of PEG solution were necessary. Each of this volume contains:

- 50 μL of MOPS buffer, 50mM pH 7.5
- 10 mg of PEG
- 250 μg biotin-PEG

After three hours of incubation at room temperature, each channel well was rinsed three times with 200 μl of PBS (0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride, pH 7.4) and, finally, slides were stored with PBS buffer at 4° C to keep them hydrated.

2.2.30 Trolox and scavenger systems preparation

50 mg of TROLOX (Hoffman-LaRoche, Switzerland) was added to 100 ml of 5 mM Tris buffer (pH 8) in a small glass placed overnight on a shaker. The absorption spectrum was measured using a Nanodrop 2000c machine, showing a peak around 290 nm (Cordes *et al.,* 2009). The solution was treated for 30 minutes with an UV lamp emitting at a wavelength of 250nm. A new peak arises at around 270 nm with an isosbestic point at 280 nm, indicating the presence of Trolox-quinone.

For the preparation of Gloxy (scavenger system) preparation, 40 µl of 34 mg/ml catalase (Roche Diagnostics, Switzerland) and 20 mg glucose oxidase type VII *from Aspergillus niger* 100,000 U/gr (Sigma-Aldrich, St. Louis, MO, USA) were added to 160 µl of T50 buffer (10 mM Tris pH8, 50 mM NaCl). After mixing, the sample was centrifuged at 14,000 rpm for 1 minute and 20 µl aliquots of the supernatant were stored at -20° C.

2.2.31 Neutravidin coating and DNA incubation
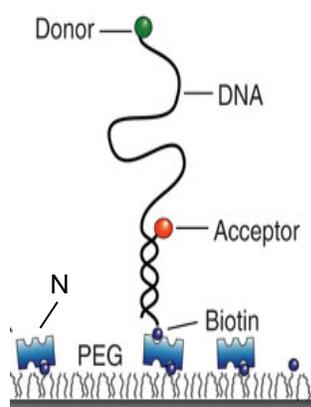
Figure 2.4: Schematic representation of DNA molecules fixation by neutravidin-biotin system. N represents neutravidin.

Neutravidin was used for immobilizing DNA molecules on PEG-coated slides (Fig. 2.4). 60 µl of neutravidin solution (0.25 mg/ml in PBS) were added to the channel after rinsing it with 300 µl of PBS. After 5 minutes, the channel was rinsed with 1 ml of PBS to remove excess neutravidin. Incubation of 25 pM biotinylated dsDNA was carried out for 5 minutes, and DNA density was checked with the emCCD camera. The channel was rinsed with 300 µl of PBS buffer to remove excess DNA. 50 µl of the following imaging buffer was added for a suitable observation of the immobilized DNA molecules:

- 25 µl KF7 buffer pH 7.5
- 23 µl Trolox 2 mM
- 1.4 µl glucose 40% (w/v)
- 0.7 µl glucose oxydase (165 U/ml) + catalase (2170 U/ml)

2.2.32 Binding assays

One nucleotide-gapped DNA (003-025-032) and both 004-025 and 004-040 DNA constructs were used as binding sensors, in which binding of a proteins is expected to induce a conformational change of the labeled DNA and thereby a change in FRET efficiency. DNA was diluted to a final concentration of 25 pM using TE buffer. After incubation of the DNA substrates, the channel well was washed with 300 µl of PBS. A volume of 50 µl of imaging buffer was added into the channel well. Table 8 summarizes different DNA substrates and concentrations used for each polymerase.

Protein	Protein concentration in the channel well	DNA substrate		
		004-025	004-040	003-025-032
Klenow Fragment	50 pM		✓	✓
	75 pM			✓
	100 pM			✓
	200 pM		✓	
	300 pM			✓
	500 pM	✓	✓	✓
	1 nM		✓	✓
	2 nM		✓	✓
HoLaMa	100 pM			✓
	300 pM			✓
	1 nM			✓
	2 nM	✓	✓	✓
	5 nM			✓
α subunit	2 nM	✓	✓	
αεθ	2 nM	✓	✓	

Table 8: List of DNA substrates and protein concentration used for each assayed polymerase (Klenow enzyme, HoLaMa, α subunit and αεθ) in smFRET binding experiments.

In the case of Klenow Fragment and HoLaMa with the 003-025-032 DNA constuct, K_D constant was calculated and the effect of complementary dTTP on HoLaMa binding to the 003-025-032 DNA construct was tested. A control sample, with 25 µl of KF7 buffer instead of polymerase was also used. This KF7 buffer was used for polymerase dilutions.

2.2.33 Real Time polymerization assay

Different polymerases (Klenow Fragment, HoLaMa, α subunit and αεθ core) were tested in order to study the changes in FRET efficiency (E^*) in real time. For this purpose, 20 pM of both 004-025 and 004-040 DNA constructs were used for initial immobilization. After successful DNA immobilization, the channel was washed with PBS before the final imaging buffer, now containing also 100 µM of each dNTP, was added. Polymerization was evaluated by taking several movies (50 seconds each) immediately after the addition of the sample into the channel well. In the case of Klenow fragment, the effect of the concentration of dNTPs during polymerization was also evaluated by using higher and lower concentrations of dNTPs (1 µM and 100 µM). The use of channels enabled us to change reaction conditions of the same sample several times.

2.2.34 TIRF analysis

Single molecule experiments were performed combining Total Internal Reflection Fluorescence (TIRF) microscopy and Alternating Laser Excitation (ALEX). TIRF microscopy allows for monitoring hundreds of single fluorescent molecules simultaneously (Moerner & Fromm, 2003). The laser light is focused into the rim of the backfocal plane of the microscope objective so that the light is finally totally reflected at the glass/water interface. An evanescent field is created in which the excitation intensity decays within a few hundred nanometers above the glass/water (sample) interface. Therefore, TIRF microscopy enables the selective illumination of surface regions and since only a small volume above the cover slide is illuminated, the background fluorescence is greatly reduced. The combination of FRET and TIRF involves the acquisition of a sequence of images of surface-immobilized molecules, followed by extraction of FRET data by image analysis. A custom-built TIRF confocal microscope with green and red lasers (Omicron, Germany) was used (Fig. 2.5), operating under the alternating-laser excitation (ALEX) (Hohlbein *et al.,* 2014).

ALEX allows the sorting of fluorescently labeled species depending on the type and quantity of fluorophores, and provides a suitable way for determining molecular distances within a heterogeneous population (Hohlbein *et al.,* 2014). ALEX consists of a rapid green and red laser switching for achieving Donor-excitation and Acceptor-excitation events for each single molecule (Kapanidis *et al.,* 2004; Hohlbein *et al.,* 2014) (Fig. 2.5). The excitation wavelengths were 561 nm and 642 nm for green and red lasers, respectively. Lasers excitation time for each frame was 50 ms. Lasers were coupled into the inverted confocal microscope by a custom-made multicolor polychroic mirror (Chroma Technology; Bellows Falls, VT, USA) while a multi-bandpass filter (Chroma Technology) blocked the remaining laser light in the emission path. The lasers intensities were set to 1.5 mW (green laser) and 0.5 mW (red laser). Laser intensities were independently controlled by a home-written LabVIEW program. The light was focused onto the back-focal-plane of the 100x NA 1.49 TIRF objective (Nikon Corporation, Tokyo, Japan) by a lens (f = 200 mm, Thorlabs) mounted on a translational stage, perpendicular to the laser beam in order to adjust the TIRF angle. After fluorescence spatial filtering by a two-lens system consisting of two tube lenses (f = 200 mm) and an adjustable slit (Thorlabs; Germany), the light is spectrally split, using dichroic mirrors (Chroma), into two beams corresponding to the green and red fluorescence detection channel. Florescence emission was imaged using an iXon Ultra 897 emCCD camera (Andor Technology; Oxford, UK).
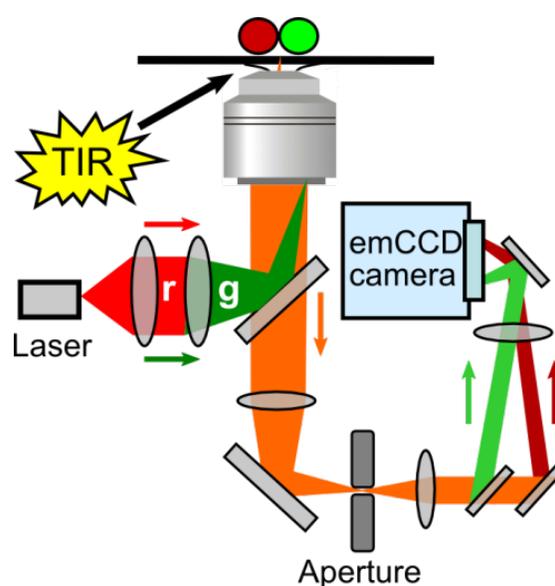
Figure 2.5: TIRF microscopy and ALEX combination for smFRET analysis (Hohlbein *et al.,* 2014).

dsDNA labeled with donor (Cy3B) and acceptor (ATTO647N) at 16-bp separation (T1B17) (Holden *et al.,* 2010) was used as a standard to check the performance of the setup. The sequence of this T1B17 standard was the following:

5' **T**AAATCTAAGTAACATAAGGTAACATAACGTAAGCTCATTCGCG 3'

3' ATTTAGATTCATTCTA**T**TCCATTGTATTGCATTCGAGTAAGCGC 5'

2.2.35 TIRF data analysis

Molecules were recorded by the emCCD camera and visualized by a Solis software for imaging (Andor Technology). TIRF data processing was developed using MATLAB, Twotone (Holden *et al.,* 2010) and TIRFDataAnalysis (Uphoff *et al.,* 2010). The software allowed generating a mapping between the donor and acceptor emission channels (transformMovie), data extraction (twotoneALEX), combine and analyse extracted data (DoItAll_TIRF.m) and change thresholds and required parameters (TirfAnalysisMain.m and DoItAll_TIRF.m). The FRET transfer efficiency (*E*) is inversely proportional to the sixth power of the distance between the two transition dipoles, where *R* is the inter-dye distance whereas R_0 is the Förster radius at which *E*=0.5 (Holden *et al.,* 2010).

$$E = 1 / [1 + (R/R_0)^6]$$

For each molecule, three photon streams (*DA, DD* and *AA*) are detected. The apparent FRET efficiency (*E**) and stoichiometry (*S*) were calculated combining several movies and including only molecules detected in both emission channels.

$$E^* = DA/(DD+DA) \qquad S = (DD+DA)/(DD+DA+AA)$$

where *DA* is the acceptor photon count after donor excitation (FRET), *DD* is the donor photon count upon direct donor excitation and *AA* is the acceptor photon count upon direct acceptor excitation (Holden *et al.,* 2010). To obtain *E*, *E** is corrected for background, direct excitation of the acceptor, leakage of the donor into the acceptor channel and the detection efficiencies of the dyes. We can obtain *S* because of the use of ALEX, since provides information about the number of

photons in the acceptor channel after direct excitation of the acceptor (*AA*) (Fig. 2.6), due to being possible to separate double-labeled molecules from donor-only and acceptor-only species. *E** provides information about donor-acceptor distance, whereas S is related to the green and the red fluorescence after green excitation to the overall fluorescence emission after green and red excitation (Hohlbein *et al.,* 2014). We can plot *E** and *S* in a two-dimensional histogram (Fig. 2.6), separating low and high FRET species and also donor-only and acceptor-only species, having an idea about the type of molecule population, specially important if we are working with a heterogeneous population and thus becoming closer to physiological conditions.
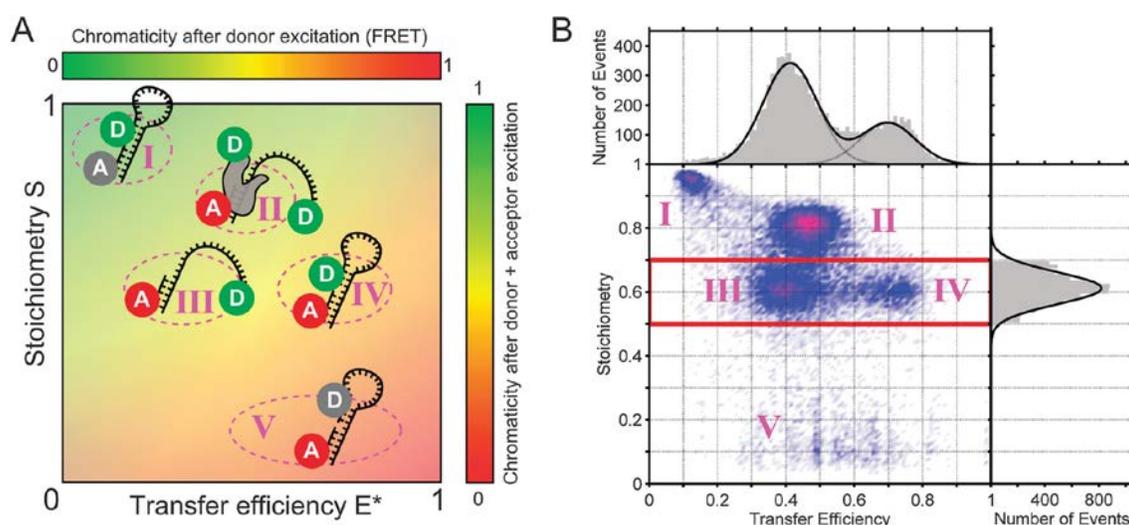
Figure 2.6: (A) Conventional single-colour laser excitation reporting *E**, which is represented as the detected chromaticity after donor excitation, from low-FRET species to high-FRET species. Only by using ALEX, the stoichiometry (*S*) enables sorting species based on their chromaticity after donor and the acceptor direct excitation: a donor-only or bleached acceptor sample (I) always with a high *S* value; an acceptor-only or a donor-bleached sample (V) with low *S* value; double-labeled molecules with an intermediate stoichiometry (II–IV). A two-dimensional *ES* histogram is generated by plotting the transfer efficiency against the stoichiometry for each single molecule, since we follow a camera-based detection system. **(B)** Two-dimensional *ES* histogram of five different molecule species diffusing through a confocal volume and using ALEX. This two-dimensional *ES* histogram obtained using ALEX enables us to understand the distribution of the different molecule species according to their labeling (different *S* values) or polymerization state (different *E** value) (Hohlbein *et al.,* 2014)

CHAPTER 3 – RESULTS

3.1 Overexpression and solubility tests

3.1.1 αH12A, αD19A and αD201A variants

The level of expression and solubility of the αH12A and αD19A proteins was comparable to that observed for α wild type. The αD201A variant, despite featuring an expression level comparable to α subunit, showed a level of solubility significantly lower, as shown in Figure 3.1, leading to experiments for improving its solubility (Fig. 3.2). Empty pBAD plasmid was used as negative control.
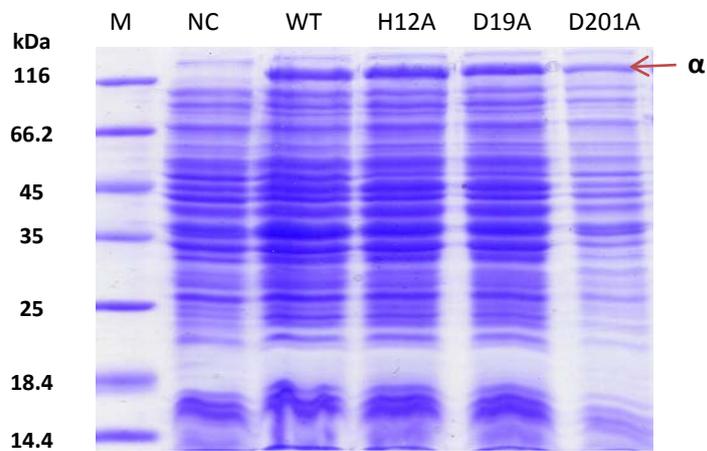
Figure 3.1: SDS-PAGE of crude extract of induced *E. coli* TOP 10 cells for overexpressing α subunit and its mutant variants. From left to right, marker (M), empty pBAD (NC), the wild type α subunit (WT), αH12A, αD19A and αD201A.
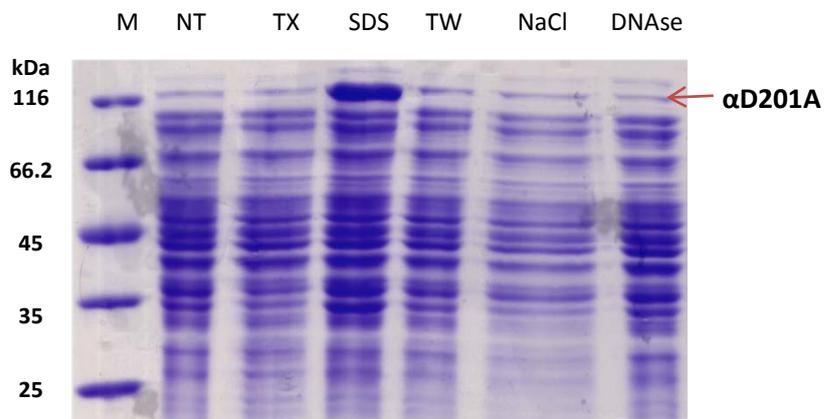
Figure 3.2: SDS-PAGE of solubilization treatments of the αD201A protein using different reagents. From left to right, marker (M), not treated sample (NT), 1% Triton X-100 (TX), 0.1% SDS, 1% Tween 20 (TW), 1M NaCl and 40 U DNAse.

Among the compounds tested to verify the eventual ability to enhance the solubility of αD201A (section 2.2.6), only 0.1% (w/v) sodium dodecyl sulfate (SDS) was able to maintain αD201A protein in the soluble fraction of the crude extract (Fig. 3.2)

3.1.2 ατεθ *complex and the αD201Aτεθ mutant variant*

As reported in Figure 3.3, the complex containing the wild type α subunit presented a slightly higher level of expression. However, no signs of insolubility or protein degradation were observed when αD201A was co-expressed with τ, ε and θ subunits, becoming hence more stable and soluble in contrast with its individual treatment.
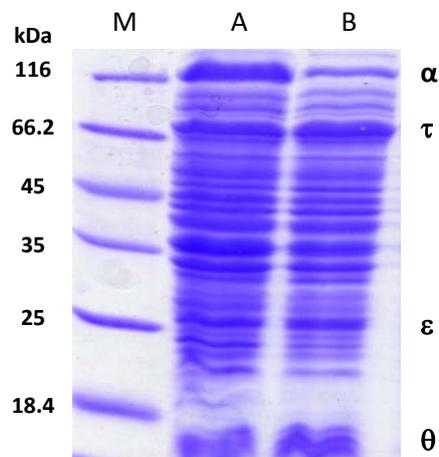
Figure 3.3: SDS-PAGE of crude extract from *E. coli* TOP 10 cells induced for the overexpression of the ατεθ (**A**) and αD201Aτεθ (**B**) complexes. M represents molecular marker.

3.1.3 HoLaMa

Protein overexpression was considerably higher when induction with 1 mM arabinose was performed for 15 hours at 30° C under constant shaking (180 rpm) in contrast with a shorter induction procedure (6 hours at 30° C) (Fig. 3.4).
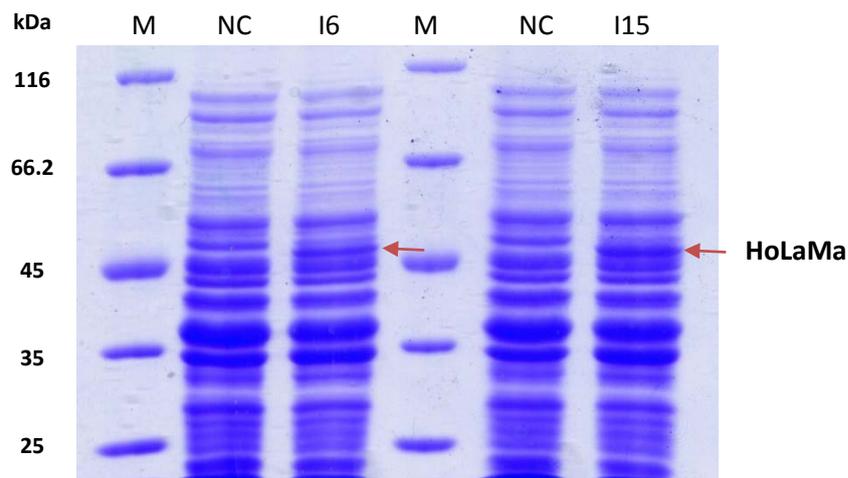
Figure 3.4: SDS-PAGE of crude extract from *E. coli* TOP 10 cells overexpressing (I) or not (NC) for HoLaMa protein after 6 or 15 hours of induction at 30° C. Empty pBAD plasmid was used as negative control. From left to right, marker (M), empty pBAD plasmid (NC), 6 hours induction (I6), marker (M), empty pBAD plasmid (NC), 15 hours induction (I15). The red harrow represents the band corresponding to HoLaMa.

3.2 Protein purification

3.2.1 DNA Pol III

3.2.1.1 Q-Sepharose anion exchange column

For each protein purification process, at least between six and twelve liters of cell culture, with resulting pellets of about 20-40 gr, were produced. Anion-exchange chromatography (Q-Sepharose FF column) was performed as the first purification step, setting a gradient of NaCl up to 600 mM for protein elution. On average, α subunit and two of its variants (H12A and D19A) eluted in a range of NaCl between 250 and 385 mM, whereas both αεθ and αD201Αεθ complexes eluted between 220 and 370 mM NaCl.

3.2.1.2 Cibacron blue affinity column

Protein elution was carried out in two steps with two different NaCl concentrations (0.15 M and 1 M). SDS-PAGE profile showed the highly-concentrated bands, corresponding to the proteins of interest, in the "*core*" and the "*tail*" fractions of the final elution peak. Only in the case of ατεθ and αD201Ατεθ complexes, the first washing step with 150 mM NaCl was omitted, due to the fact that a high amount of complex was lost in preliminary studies during that washing step, lowering protein yield after purification.

3.2.1.3 HiTrap Heparin affinity column

With the aim of improving the purity of the recovered proteins and to remove, as well, the RNA polymerase, a HiTrap Heparin HP affinity column step was performed. Once recovered, the "*core*" and the "*tail*" fractions from the previous purification step, the resulting sample was conditioned to eliminate $MgCl_2$ and bring the concentration of NaCl to 50 mM. After loading the protein onto the column, protein elution took part with a gradient of NaCl up to 300 mM at 1 ml/min. Single α polymerase as well as H12A and D19A variants eluted between 175 and 228 mM NaCl. On the other hand, ατεθ and αD201Ατεθ complexes elution occurred in a range of NaCl between 195 and 290 mM.
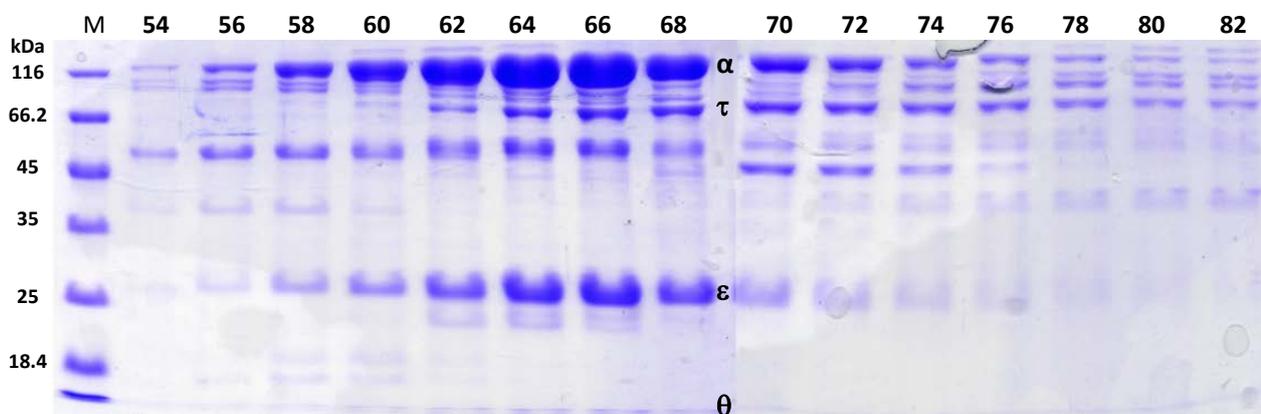
Figure 3.5: SDS-PAGE of complex after HiTrap Heparin purification step. From left to right, marker (M) and fractions from 54 to 82.

3.2.1.4 Hitrap Q anion exchange column

This purification step was exclusive for the α subunit and the αεθ, since both samples presented the highest yield of purified protein. Those fractions eluted from the Hitrap Heparin affinity column with the highest concentration of proteins of interest were pooled and concentrated with Amicon Ultrafiltration Cells, and subjected to this anion exchange chromatography with a NaCl gradient up to 600 mM, reached in 50 ml. Protein elution occurred at the end of the gradient. From SDS-PAGE analysis, it was clear that proteins were consistently purified (Fig. 3.6).
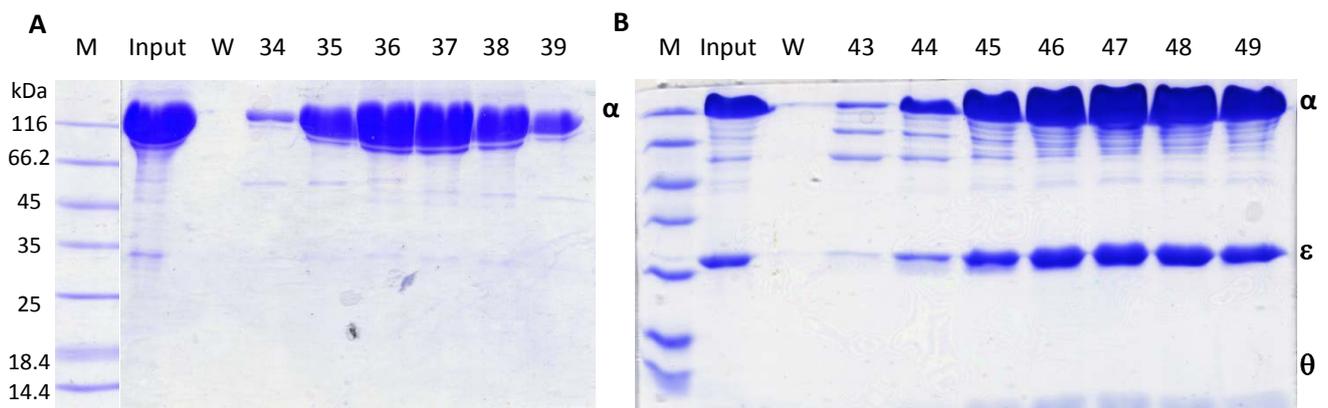
Figure 3.6: (A) SDS-PAGE of α subunit after HiTrap Q purification step. From left to right, marker (M), input, 150 mM NaCl washing step (W) and fractions 34 to 39. **(B)** SDS-PAGE of αεθ after HiTrap Q purification step. From left to right, marker (M), input, 150 mM NaCl washing step (W) and fractions 43 to 49.

3.2.1.5 Sephacryl S-300 Gel Filtration column

This final purification step was performed exclusively at the end of the purification of both αεθ and αD201Αεθ complexes, so as to determine their own identity and the type of protein sub-assemblies present in the purified complexes.

Fractions of interest eluted from the HiTrap Heparin column were loaded onto a Sephacryl S-300 column using a loop of 2 ml. The elution of both complexes occurred at the same volume as that observed for thyroglobulin (about 50 ml). Nevertheless, an excess of αεθ was also observed and confirmed by SDS-PAGE (Fig. 3.8). This second peak was significantly higher in the case of the αD201Αεθ complex, instead of the αεθ complex where both peaks were similar (Fig. 3.7). Surprisingly, a peak corresponding to dimeric association was not observed.
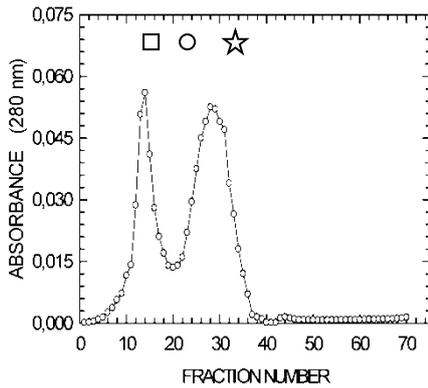
Figure 3.7: Sephacryl S-300 column of the trimeric replicase as final purification step. The square, circle and star indicate the elution volume of the molecular mass markers thyroglobulin (669 kDa), ferritin (440 kDa), and catalase (232 kDa), respectively. The theoretical molecular masses of ατεθ and αεθ complexes are equal to 711 and 166 kDa, respectively.

A   **B**
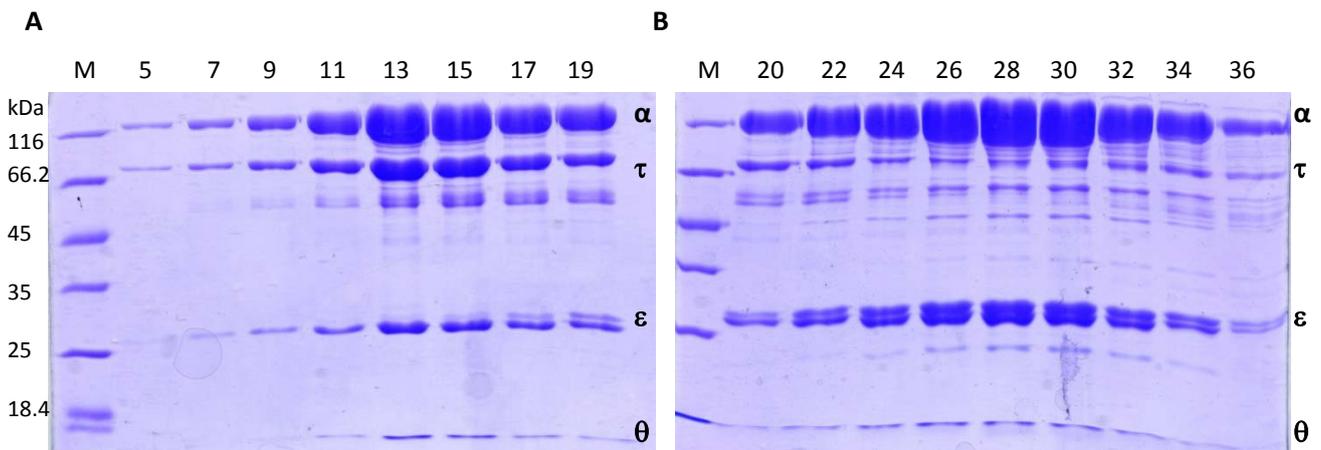
Figure 3.8 SDS-PAGE of ατεθ complex after Sephacryl S-300 (gel filtration) column. **(A)** Fractions from the peak representing the trimeric *in vivo* sub-assembly (5-19); **(B)** Fractions representing the excess αεθ (20-36).

3.2.2 HoLaMa

3.2.2.1 Q-Sepharose anion exchange column

The crude protein extract was loaded onto a Q-Sepharose FF anion exchange column, and a NaCl gradient up to 600 mM after 400 mL was applied. HoLaMa eluted during the washing of the column with equilibration buffer (50 mM NaCl).

3.2.2.2 Cibacron blue affinity column

The sample collected with the washing step from Q-Sepharose column, containing HoLaMa, was concentrated using Amicon Ultrafiltration Cells fitted with an YM-30 membrane. Sample was then loaded onto the Cibacron Blue affinity column. A single washing step with buffer containing 1 M NaCl was performed, reaching a peak in which "*head*", "*core*" and "*tail*" fractions were collected. Protein was mainly eluted in the "*core*" and "*tail*" fractions.

3.2.2.3 Hitrap Heparin affinity column

Following the Cibacron blue purification step and protein concentration by Amicon Ultrafiltration Cells, the sample was loaded onto a HiTrap-Heparin column. Protein elution was effected by a NaCl gradient up to 300 mM, obtaining considerably pure eluted fractions, as shown in Fig. 3.9. These fractions were pooled and subsequently concentrated up to a final volume of 2 mL using Amicon Ultrafiltration Cells fitted with a YM-30 membrane. This final sample was used for testing enzymatic activity.
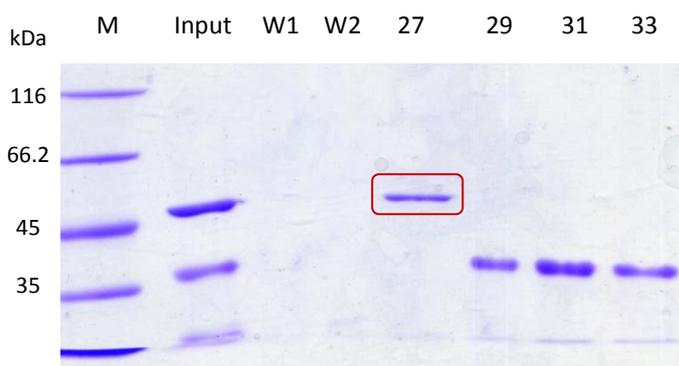
Figure 3.9: SDS-PAGE of HoLaMa after HiTrap Heparin purification step. From left to right, marker (M), Input, 150 mM NaCl washing step (W1), 1M NaCl washing step (W2) and fraction 27 to 33. HoLaMa is contained in fraction 27, highlighted in red.

3.3 Cross-linking experiments and mass spectrometry analysis

The molecular mass of the eluted complex was determined by electrospray ionization mass spectrometry (ESI-MS). Results revealed that the mass of the complex was 708.05 kDa (Fig. 3.10), meaning that a trimeric complex, sub-assembled *in vivo*, was obtained. No dimeric complex was present, since we did not observed any peak at 474.2 kDa (Fig. 3.10.B). The observed peak at 512 kDa could represent degraded trimeric complex.

In addition, interactions between subunits were identified by subjecting the trimeric complex to cross-linking techniques, using bis (sulfosuccinimidyl) suberate (BS3) (Fig. 3.11). Mass spectra of the cross-linked trimeric polymerase revealed the presence of α-ε, ε-θ, τ-τ, and α-α contacts. Table 9 shows cross-linked peptides identified with MS analysis of in gel trypsin-digested covalent complexes. Tables 23, 24, 25, and 26 (Appendix I) represent mass spectrometry data obtained with in gel digested proteins leading to the identification of α, τ, ε and θ subunits, respectively.
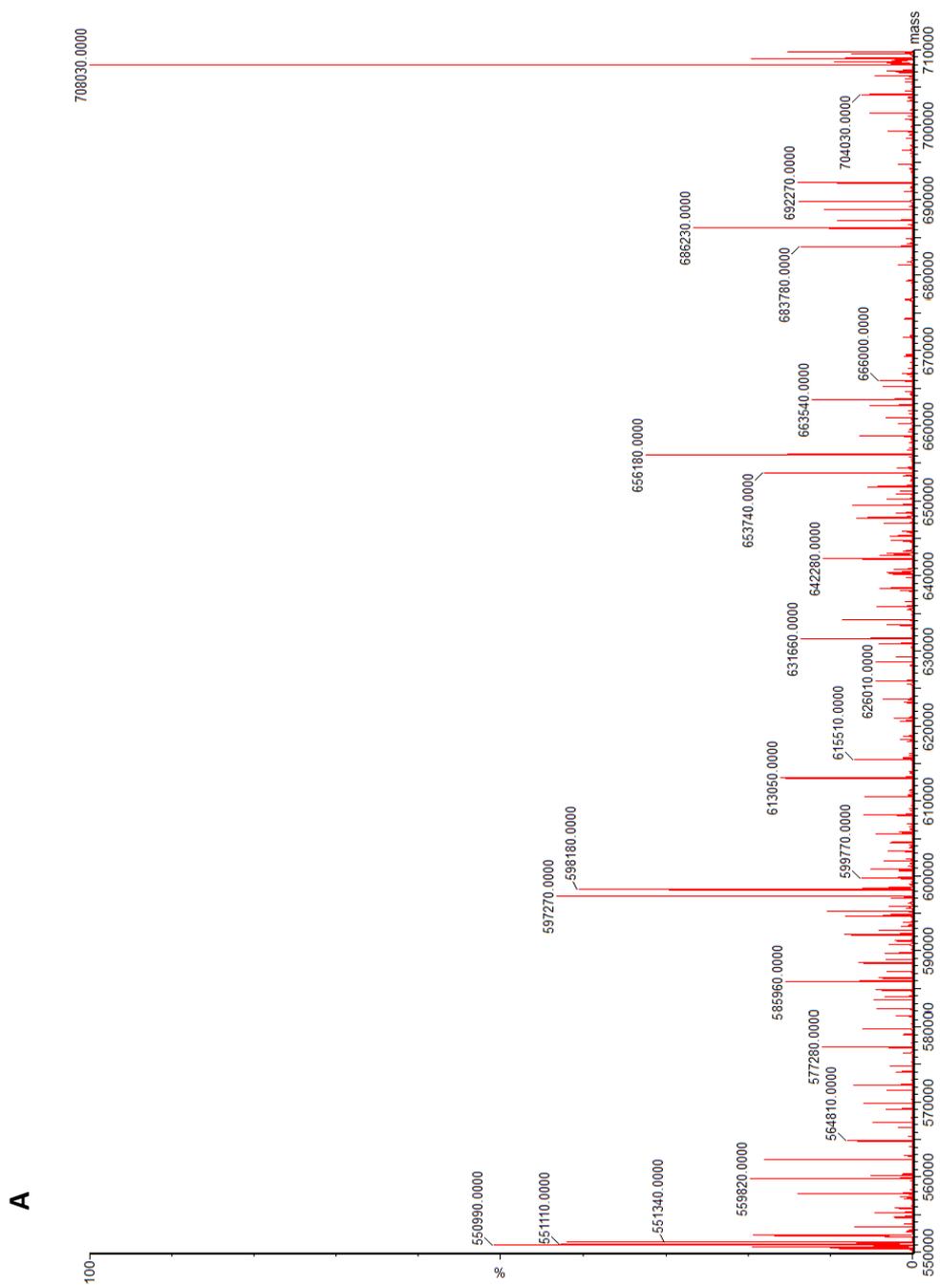
Figure 3.10.A: ESI-MS of the trimeric replicase. The peak on the right at 708.050 kDa represented the trimeric sub-assembly.
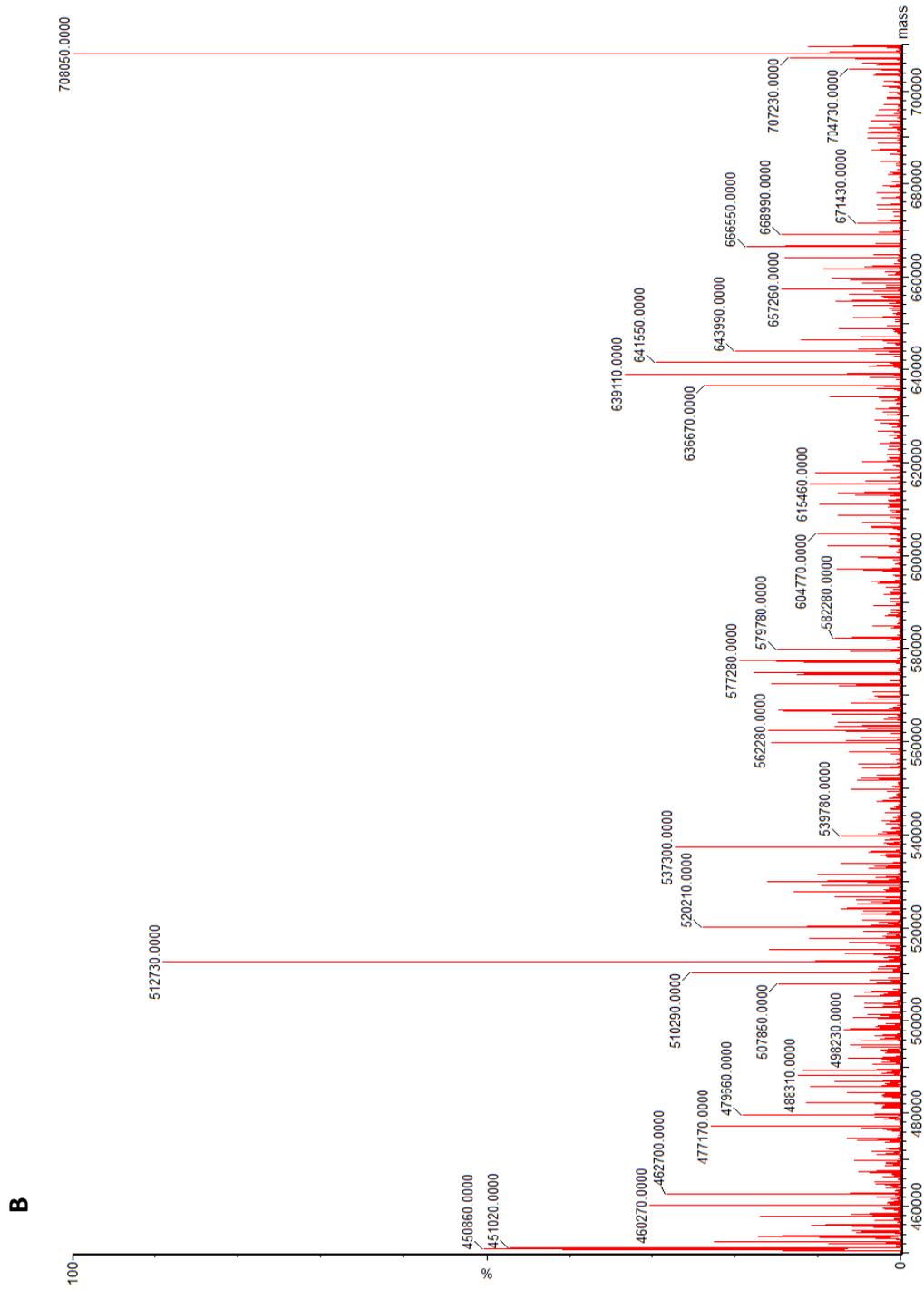
Figure 3.10.B: Once again, the peak at 708.050 kDa represented the trimeric assembly. No dimeric assembly was present, since there was not peak at 474.2 kDa. A peak at 512 kDa may represent degraded trimeric αεθcomplex.

Figure 3.11: SDS-PAGE of purified αεθ complex subjected or not to cross-linking with BS3. The trimeric replicase (350 nM) was incubated with 5, 50, 500, or 1000 μM BS3 in PBS (lanes II-V, respectively), for 1 h at room temperature. Reactions were blocked by the addition of 100 mM Tris-HCl, pH 7.9. Lane I was loaded with an aliquot of untreated αεθ complex. M represents the molecular marker

Cross-linked peptide	Expected M.W.	Theoretic M.W.	Involved residues
α-α interaction			
[1025-1046 (α)]-[1025-1046 (α)]	5584.932	5584.932	K1037(α) and/or K1044 (α)
α-ε interaction			
[217-232 (α)]-[205-213 (ε)]	3043.705	3043.704	K229 (α) and K211 (ε)
[242-266 (α)]-[229-243 (ε)]	4732.356	34732.355	K234 (α) or K235 (α) and K211 (ε)
ε-θ interaction			
[43-56 (ε)]-[4-15 (θ)]	3215.569	3215.568	K53 (ε) and K7 (θ)
[43-56 (ε)]-[8-28 (θ)]	4045.002	4045.001	K53 (ε) and K15 (θ)
τ-τ interaction			
[457-477 (τ)]-[457-477 (τ)]	4754.671	4754.670	K468 (τ) and/or K472 (τ) and/or K473 (τ)

Table 9: Cross-linked peptides identified with ESI-MS analysis of *in gel* trypsin-digested covalent complexes

3.4 Enzymatic activity assays in microplate

The crude protein extract of the ατεθ complex, obtained after sonication of 3 liters of cell cultures, was loaded onto a Sephacryl S-300 gel filtration column. The resulting fractions were subjected to enzymatic activity assays, specific for each subunit.

As noticed from the three chromatograms of polymerase, ATPase and exonuclease activities, shown in Figures 3.12.A, B and C, respectively, results confirmed the presence of a peak of activity in fractions 13 and 14, corresponding to the elution volume of thyroglobulin (660 kDa), consistent with a spontaneous tendency of a trimeric sub-assembly of the DNA pol III *in vivo,* further discussed in the following chapter.
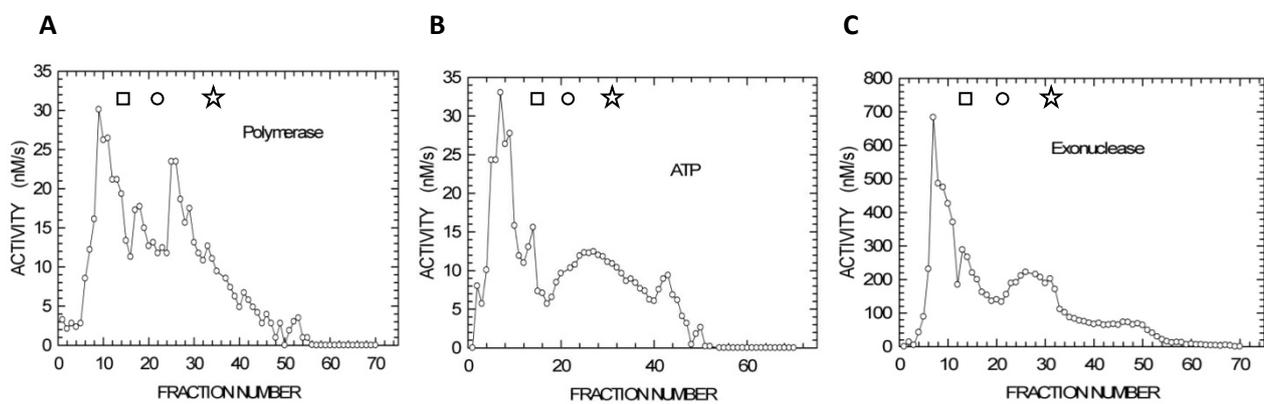
Figure 3.12: Polymerase (**A**), ATPase (**B**) and 3' → 5' exonuclease (**C**) activities detected in the fractions eluted from the Sephacryl S-300 column. The square, circle, and star indicate the elution volume of the molecular mass markers thyroglobulin (669 kDa), ferritin (440 kDa), and catalase (232 kDa). The theoretical molecular masses of ατεθ and αεθ are equal to 711 and 166 kDa, respectively.

3.5 Enzyme kinetics by standard enzymatic assays

3.5.1 DNA Pol III

3.5.1.1 Pyrophosphatase activity assay

The pyrophosphatase activity measurements were carried out in the presence of 1 mM sodium pyrophosphate as substrate. Under these conditions, the wild type α subunit reached a pyrophosphatase activity equal to 21 nM/s. In contrast, both αH12A and αD19A mutant proteins showed a significantly impaired activity, 10 times (2.1 nM/s) and 5.5 times (3.8 nM/s) lower, respectively (Fig. 3.13.A).

On the other hand, the αD201τεθ complex presented an activity of pyrophosphate hydrolysis almost six times higher (63 nM/s) than that of the wild type ατεθ complex (11 nM/s) (Fig. 3.13.B).
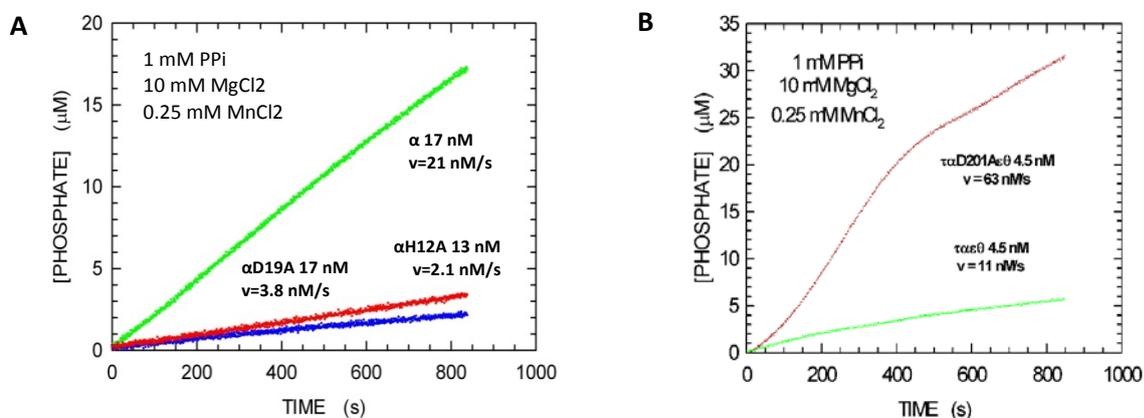
Figure 3.13: Pyrophosphatase activity using 1 mM inorganic pyrophosphate as substrate. **(A)** Pyrophosphatase activity of the wild type α subunit (green dots), the αD19A (red dots) and the αH12A (blue dots); **(B)** Pyrophosphatase activity of the ατεθ complex (green dots) and the αD201Aτεθ complex (red dots).

3.5.1.2 Pyrophosphatase activity inhibition assay

The pyrophosphatase activity of α subunit was inhibited by sodium fluoride (NaF). In the absence of fluoride, 30 nM α subunit showed pyrophosphatase activity equal to 17.8 nM/s. However, in the presence of NaF, the pyrophosphatase activity of α subunit was strongly inhibited at different concentrations of sodium fluoride (Fig. 3.14.B). Inorganic pyrophosphatase (0.45 nM) was used as control and its intrinsic activity (47 nM/s) was also highly inhibited in the presence of NaF (Fig. 3.14.A).
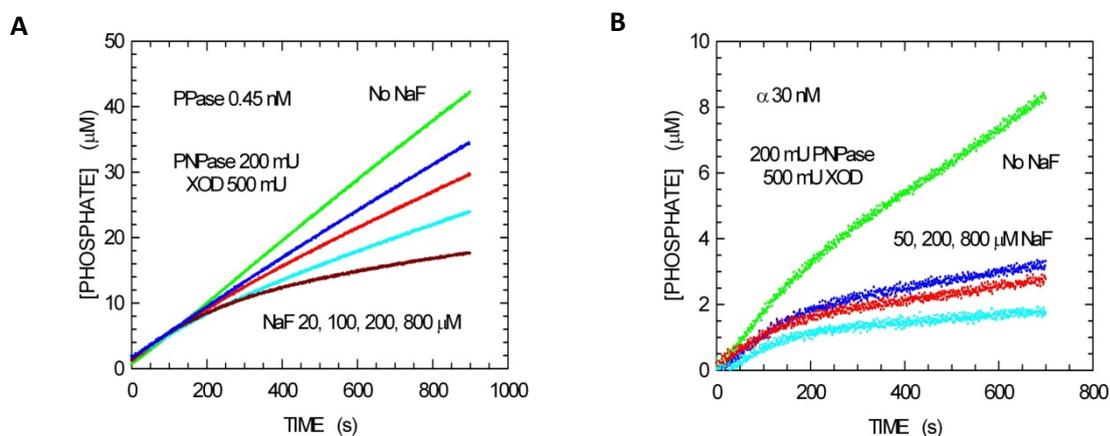
Figure 3.14: Inhibition of the pyrophosphatase activity of the inorganic PPase **(A)** and the wild type α subunit **(B)** by different concentrations of sodium fluoride (NaF).

3.5.1.3 Polymerase activity

When 17 nM of the wild type α subunit was tested in the presence of 1 μM DNA, reaction velocities were equal to 24.8 and 17.3 nM/s in the presence and absence of inorganic PPase, respectively. On the other hand, the polymerization rate for H12A was 4.9 nM/s in the presence of inorganic PPase, and 1.8 nM/s in the absence of an excess of PPase. The activity rates for the D19A variant were 8 and 2.9 nM/s in the presence and absence of inorganic PPase, respectively (Fig. 3.15 A and B).
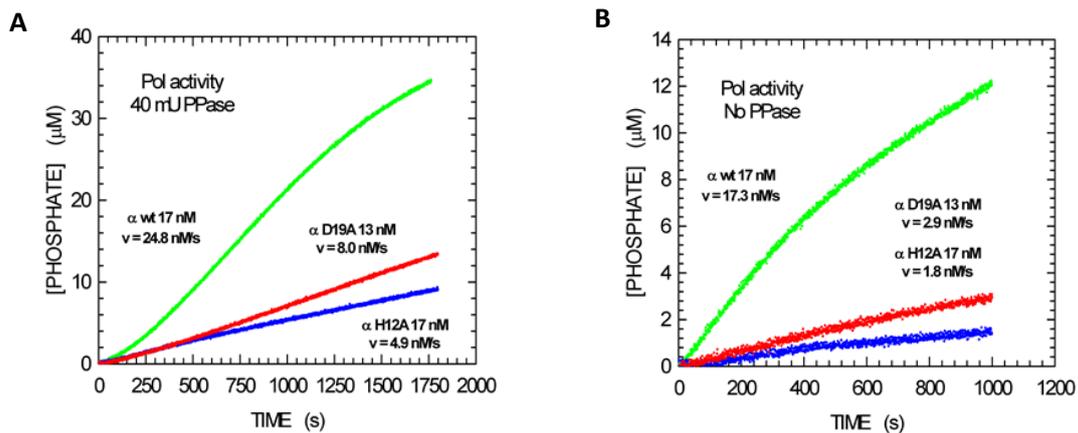
Figure 3.15: Polymerase activity of the wild type α subunit (green dots), the D19A (red dots) and H12A (blue dots) in the presence (**A**) and in the absence (**B**) of exogenous PPase.

The trimeric ατεθ complex (5 nM) showed a polymerase reaction rate equivalent to 19.9 and 14.7 nM/s in the presence and in the absence of inorganic PPase, respectively. Finally, the corresponding velocities for the αD201Aτεθ complex (3.2 nM) in the presence and in the absence of exogenous inorganic PPase were 24.5 and 13.2 nM/s, respectively (Fig. 3.16 A and B).
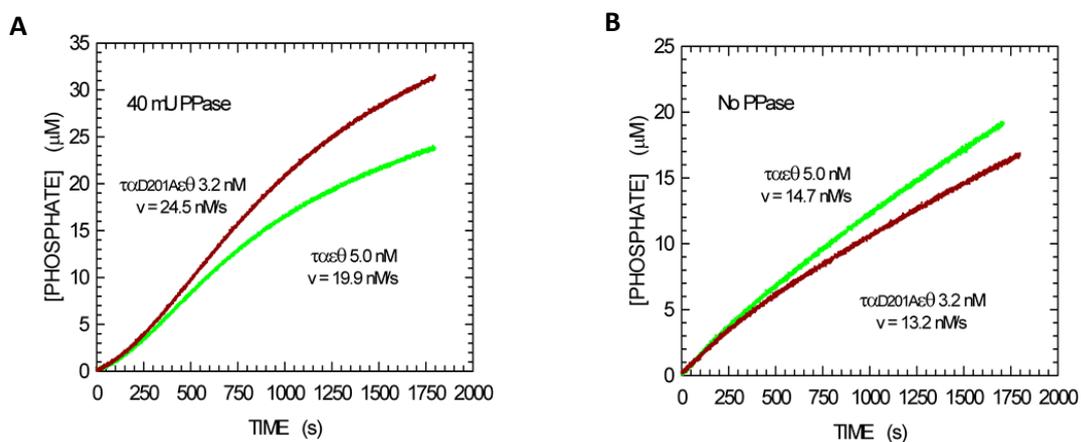
Figure 3.16: Polymerase activity of the wild type α complex (green dots) and the variant (red dots) in the presence (**A**) and in the absence (**B**) of exogenous PPase.
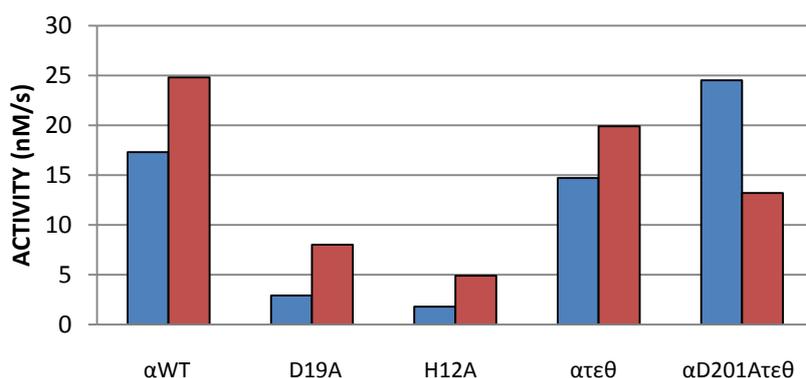
Figure 3.17: Graphic representing polymerase activity in the absence (blue columns) and in the presence (red columns) of exogenous PPase for the wild type α subunit, the αD19A and αH12A mutant variants, the wild type ατεθ complex and the αD201Αατεθ complex.

3.5.1.4 Study of the dependence of inter-primer distance of double-primed DNAs on polymerase activity

Steady-state assays were also performed for studying the effect of inter-primer distance on DNA replication, using α subunit (92 nM) and a single substrate (100mer, 75mer, 50mer or 25mer DNA) in the presence of 100 µM dNTP. Under these conditions, the highest DNA polymerase activity was observed for the 25mer DNA (Fig. 3.18; fittings in Appendix II). Besides, when the double-primed DNAs (dpDNA) were evaluated (sequences in section 2.2.21), the template head, which depends on dATP for DNA elongation, was replicated almost at the same initial velocity as the tail, whose elongation requires dTTP, using the 75mer or the 100mer DNAs; however, a slower rate of replication was observed in the presence of the 50mer DNA (Fig. 3.18; fittings in Appendix II). Data regarding initial velocities for dpDNA replication under these conditions are presented in Table 10.

dNTP (100 µM)	25mer	50mer	75mer	100mer
dTTP	-	1.7 nM/s	2.2 nM/s	2.2 nM/s
dATP	-	3.3 nM/s	2.5 nM/s	2.9 nM/s
dGTP	13.6 nM/s	-	-	-

Table 10: Initial velocities for α subunit (92 nM) using 25mer, 50mer, 75mer or 100mer dpDNA substrate in the presence of 100 µM dTTP, dATP or dGTP.

97
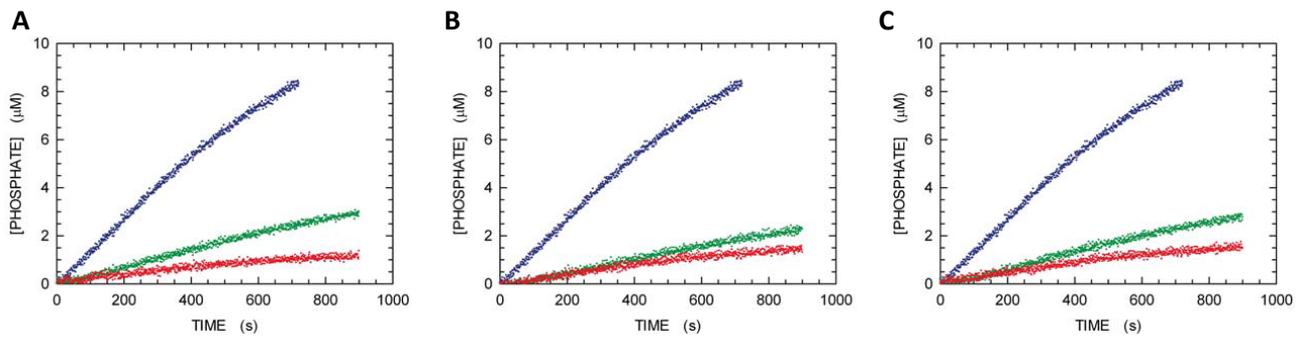
Figure 3.18: Steady-state assays of 92 nM α subunit DNA polymerase activity using 1 µM DNA and 100 µM dNTP. **(A)** Kinetics observed in the presence of 25mer DNA and dGTP (blue dots, also reported in panels **B** and **C**), 50mer DNA and dTTP (red dots), or 50mer DNA and dATP (green dots). **(B and C)** Reactions assayed in the presence of dTTP (red dots) or dATP (green dots), using 75mer (**B**) or 100mer (**C**) DNA.

3.5.1.5 Single turnover assays with the trimeric ατεθ complex

Single-turnover assays were developed using the trimeric ατεθ replicase (0.2 µM), a double-primed DNA (0.2 µM) and the 25mer DNA (0.2 µM), in the presence of 100 µM dNTPs. In paragraph 2.2.21, sequences of DNAs used in single-turnover assays are shown. dTTP is required for the extension near the 3' end of the template (head), whereas dATP is necessary for the extension near the 5' end of the template (tail). dGTP is the only required nucleotide for the elongation of the single-primed DNA. The double-primed DNA was used as a model of lagging strand whereas the single-primed DNA mimicked the leading strand. Initially, single-turnover assays were developed in the presence of 100 µM dTTP, 200 nM 25mer DNA, and 200 nM of 100mer, 75mer, or 50mer DNA, with the purpose of selecting the most suitable substrate for evaluating DNA polymerase activity of the ατεθ complex (200 nM). Under these conditions, the extension velocity of the 100mer was considerably higher (9 nM/s) than those of 75mer (1.86 nM/s) and 50mer (1.25 nM/s) (Fig. 3.19).
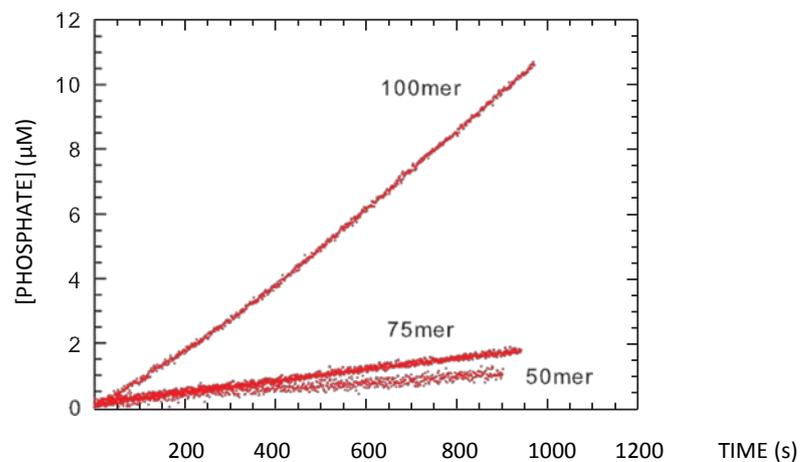
Figure 3.19: Single-turnover assays of DNA polymerase activity of 200 nM ατεθ complex, in the presence of 100 µM dTTP, 200 nM 25mer DNA, and 200 nM of 100 mer, 75mer or 50mer DNA.

Once selected the 100mer as the best alternative, the extension velocity of the 100mer DNA at the head site in the presence of dTTP was 9 nM/s. When both dTTP and dATP were present for replicating the double-primed 100mer DNA head and tail, polymerase activity was equal to 19 nM/s. Finally, in the presence of dTTP, dATP, and dGTP (100 μM each), the extension rate increased up to 38 nM/s (Fig. 3.20). The corresponding activity for the 25mer DNA was equivalent to 19 nM/s, twice the activities observed with the double-primed DNA.
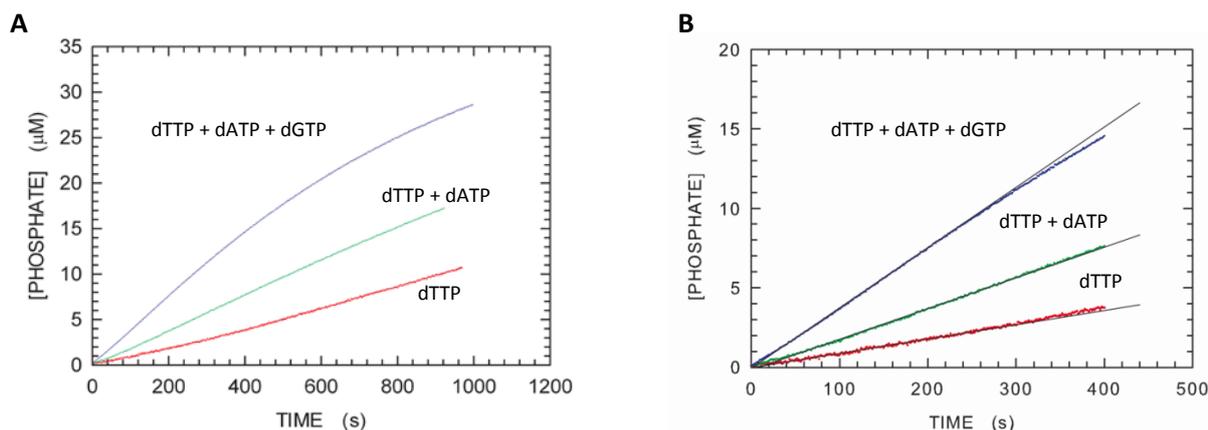
Figure 3.20: (A) Single-turnover assays of DNA polymerase activity of 200 nM of the trimeric αεθ complex, in the presence of 25mer and 100mer DNA (200 nM each). Reaction mixtures contained dTTP (100 μM, red dots), dTTP and dATP (100 μM each, green dots), or dTTP, dATP and dGTP (100 μM each, blue dots). **(B)** Detail of the reaction kinetics reported in Fig. 3.20.A. Continuous lines represent the best fits used to calculate initial reaction velocities.

3.5.1.6 Exonuclease activity assay

The exonuclease activity assay showed that the αεθ complex (1.8 nM) catalyzed this reaction at a speed equivalent to 42.5 nM/s; the exonuclease speed of αεθ was 35.5 nM/s, using a concentration three times higher (5.3 nM). The same test also showed how α subunit lacked the exonuclease activity (Fig. 3.21).
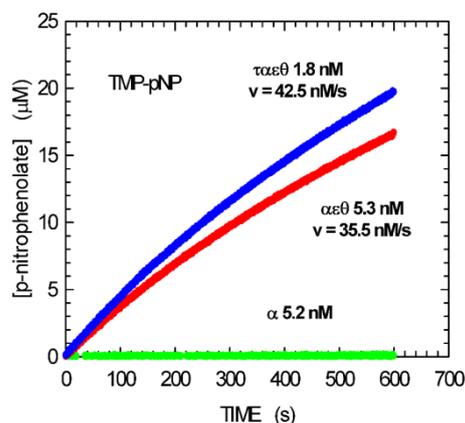
Figure 3.21: 3' → 5' exonuclease activity of the αεθ complex (blue dots), αεθ (red dots) and the wild type α subunit (green dots) by measuring the cleavage of the TMP-pNP substrate (Hamdan *et al.,* 2002b).

3.5.2 HoLaMa

3.5.2.1 Polymerase activity

Different substrate concentrations were used to evaluate the rate of reaction of Klenow enzyme (control) and HoLaMa. Steady state measurements were carried out in the presence of 100 µM dTTP with a double-stranded DNA that allows the extension of 25 nucleotides and resulted from the annealing of a 40mer DNA template and a 15mer DNA primer. Table 11 summarizes polymerization rates for Klenow enzyme and HoLaMa at different DNA concentrations. Results are further discussed in chapter 4.3 (Fig. 4.9).

[DNA] (nM)	Polymerization speed (nM/s)	
	Klenow enzyme (1.5 nM)	HoLaMa (51 nM)
100	4.2	1.8
200	9.8	2.7
500	18.4	7
750	23.2	9.2
1000	26.3	10.75
1500	30.8	14.2

Table 11: Polymerization velocities for Klenow enzyme (1.5 nM) and HoLaMa (51 nM) under steady state conditions using a 40mer DNA with 25 extending nucleotides, in the presence of 100 µM dNTPs.

3.5.2.2 Polymerase activity under non-processive conditions

The polymerase assay under non-processive conditions was developed by using a DNA from the annealing of a 16mer template and a 15mer primer, thus allowing the extension of a single nucleotide and making valuable the efficiency of catalysis in the absence of processivity. Klenow enzyme showed an initial polymerase velocity equal to 3 nM/s, whereas HoLaMa showed an initial enzymatic activity equivalent to 9.4 nM/s (Fig. 4.10, chapter 4).

3.6 Growth kinetics and phenotypes

As previously described in paragraph 2.2.17, bacterial growth curves of induced and non-induced α wild type protein and two of its mutant variants (H12A and D201A) were determined. These analyses were performed in order to conclude whether the mutations affected or not cell growth and/or were involved in the development of a different phenotype.

As reported in figure 3.22 A and B, a gap between the curves of the induced and non-induced strains was found in both mutant variants, although the αD201 mutant variant showed the most significant deviation between the induced and non-induced cultures. Moreover, the induced αD201 mutant displayed the lowest final bacterial biomass (Fig. 3.22.A). Not significant differences were found in terms of final growth between the curve of the induced α wild type strain and that of induced αH12A strain (Fig. 3.22.B). The effect related to the growth of these site-specific mutant strains was also evident in the case of the non-induced cultures, especially for the αD201 variant, which was extremely more stressed by the effect of bearing the mutation, in comparison with the observations in the αH12A strain. Non-induced αD201A did not achieve the same growing rate than that of α wild type, reaching a considerably weaker final biomass (Fig. 3.22).
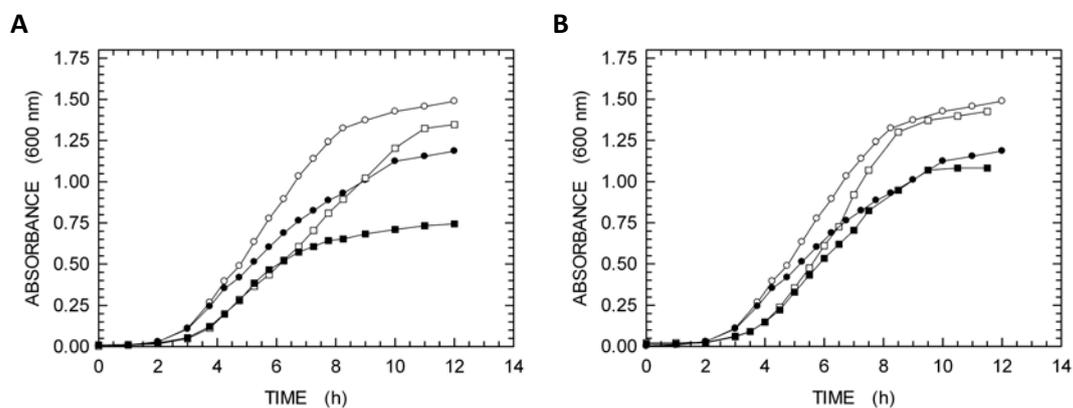
Figure 3.22: Growth kinetics of *E. coli* overexpressing (filled symbols) or not (empty symbols) the wild type α subunit (circles), the D201A (squares, **A**) or the H12A (squares, **B**) mutant variants.

Differences related to cell growth among diverse bacterial strains were also confirmed by calculating population density. The title of all strains was determined by plating 100 µl of each culture on LB dishes, antibiotic free, at fixed times: 0, 4, 4:40, 5:20, 6, 8 and 10:30 hours after induction. Table 12A summarizes the comparison of population densities of *E. coli* overexpressing or not the wild type subunit and the αD201A variant, whereas Table 12B represents the comparison between the wild type protein and the

αH12A strain. αD201A presented a remarkably lower population density, in contrast with α wild type. Also αH12A showed a decrease of population density, but not as significant as in the case of the αD201A variant. However, the mortality after eight hours was indeed higher in the case of the αH12A (Fig. 3.23.B).

Time	Wild type strain	Population density	Mutant strain	Population density
0 hours	αNI	2.89×10^{10} ufc/ml	αD201A NI	3.44×10^{10} ufc/ml
4 hours	αNI	1.5×10^{7} ufc/ml	αD201A NI	8.68×10^{6} ufc/ml
	αIND	2.38×10^{6} ufc/ml	αD201A IND	2.79×10^{6} ufc/ml
4:40 hours	αNI	4.13×10^{7} ufc/ml	αD201A NI	5.35×10^{6} ufc/ml
	αIND	4.1×10^{6} ufc/ml	αD201A IND	4.83×10^{6} ufc/ml
5:20 hours	αNI	1.29×10^{8} ufc/ml	αD201A NI	6.66×10^{7} ufc/ml
	αIND	2.62×10^{7} ufc/ml	αD201A IND	4.55×10^{7} ufc/ml
6 hours	αNI	2.95×10^{8} ufc/ml	αD201A NI	1.48×10^{8} ufc/ml
	αIND	1.44×10^{8} ufc/ml	αD201A IND	1.3×10^{7} ufc/ml
8 hours	αNI	5.97×10^{9} ufc/ml	αD201A NI	5.35×10^{8} ufc/ml
	αIND	5.75×10^{8} ufc/ml	αD201A IND	9×10^{7} ufc/ml
10 hours	αNI	1.25×10^{9} ufc/ml	αD201A NI	3.05×10^{8} ufc/ml
	αIND	3.7×10^{8} ufc/ml	αD201A IND	3.12×10^{7} ufc/ml

Table 12A: Growth kinetics of induced and non-induced *E. coli* cells hosting the wild type α subunit and the αD201A variant, growing at 37° C.

Time	Wild type strain	Population density	Mutant strain	Population density
0 hours	αNI	2.89×10^{10} ufc/ml	αH12A NI	$1,89 \times 10^{10}$ ufc/ml
4 hours	αNI	1.5×10^{7} ufc/ml	αH12A NI	$9,65 \times 10^{6}$ ufc/ml
	αIND	2.38×10^{6} ufc/ml	αH12A IND	$1,56 \times 10^{7}$ ufc/ml
4:40 hours	αNI	4.13×10^{7} ufc/ml	αH12A NI	$4,45 \times 10^{7}$ ufc/ml
	αIND	4.1×10^{6} ufc/ml	αH12A IND	$6,6 \times 10^{6}$ ufc/ml
5:20 hours	αNI	1.29×10^{8} ufc/ml	αH12A NI	$9,5 \times 10^{7}$ ufc/ml
	αIND	2.62×10^{7} ufc/ml	αH12A IND	$7,35 \times 10^{6}$ ufc/ml
6 hours	αNI	2.95×10^{8} ufc/ml	αH12A NI	$1,56 \times 10^{8}$ ufc/ml
	αIND	1.44×10^{8} ufc/ml	αH12A IND	$2,8 \times 10^{7}$ ufc/ml
8 hours	αNI	5.97×10^{9} ufc/ml	αH12A NI	$3,57 \times 10^{9}$ ufc/ml
	αIND	5.75×10^{8} ufc/ml	αH12A IND	$9,82 \times 10^{8}$ ufc/ml
10 hours	αNI	1.25×10^{9} ufc/ml	αH12A NI	$1,2 \times 10^{8}$ ufc/ml
	αIND	3.7×10^{8} ufc/ml	αH12A IND	9×10^{6} ufc/ml

Table 12B: Growth kinetics of induced and non-induced *E. coli* cells hosting the wild type α subunit and the αH12A variant, growing at 37° C.
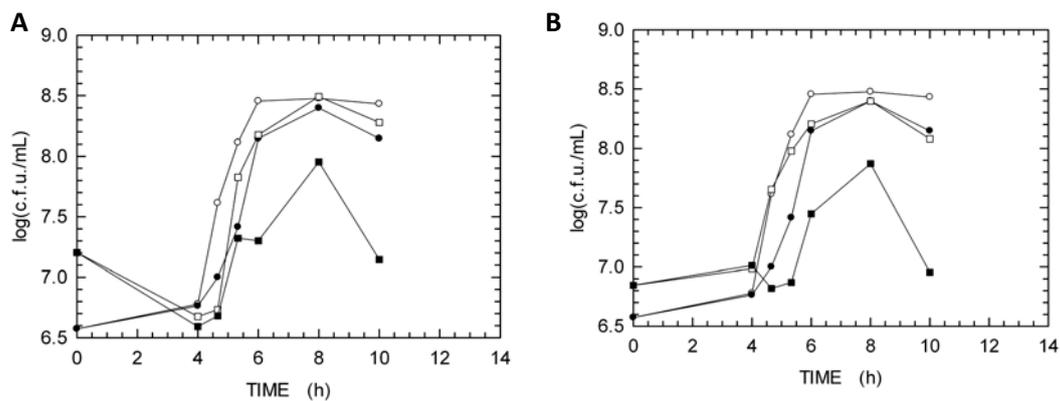
Figure 3.23: Population density of *E. coli* overexpressing (filled symbols) or not (empty symbols) the wild type α subunit (circles), the D201A (squares, **A**) or the H12A (squares, **B**) mutant variants.

In addition, phenotype modifications conferred by the overexpression of the αH12A or αD201 mutants, compared to that of observed following the overexpression of the wild type α protein, were also analyzed. During bacterial growth, strains bearing α subunit or two of the mutant variants (D201A and H12A) were subjected to phenotypic examination, using a fluorescence microscope and Hoechst dye. *E. coli* cells bearing the wild type α polymerase were normal-sized, rod-shaped cells and presented the typical motility and displacement (Fig. 3.24). In contrast, bacterial cells bearing the H12A and, especially, the D201A mutation were visibly longer and practically motionless. The effect of mutation was higher for the D201A variant, for which the presence of cell nucleoids, caused by the inability to achieve the complete cell division by binary fission, was also noted (Fig. 3.24). The decrease of final biomass was again confirmed.

Finally, mutator phenotype assays were performed using LB dishes, rifampicin supplemented (0.5 mg/ml), at 4, 6 and 10 hours after the induction of the wild type α subunit, the αD201A or the αH12A. Plates were incubated for 24 hours at 37° C and the average number of single colonies present in three serial dilutions was calculated. *E. coli* populations induced or not to overexpress the α wild type subunit presented a frequency of mutation equal to $6.7 \pm 5.9 \times 10^{-8}$ and $1.4 \pm 0.4 \times 10^{-8}$, respectively. On the contrary, when both non-induced and induced cultures of *E. coli* hosting the D201A or the H12A variant were assayed, the mutation frequency after 10 hours culturing at 37° C reached an order of magnitude equivalent to 10^{-2}.
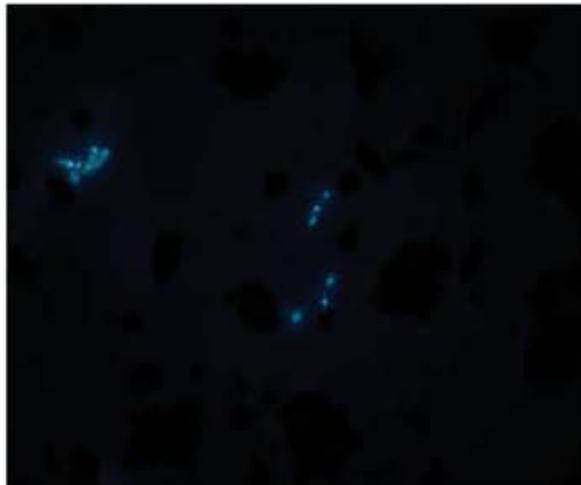
Figure 3.24: Fluorescence microscopy representing different phenotypes of *E. coli* cells overexpressing the wild type α subunit (top panel) or the D201A mutant variant (middle and bottom panel).

3.7 smFRET experiments

3.7.1 Endpoint assays for DNA polymerization

The endpoint assay was carried out for screening for the maximum FRET change between the initial state (long ssDNA overhang) and the final state (dsDNA). For each endpoint assay, 10 nM of each DNA construct, 100 µM dNTPs and 40 nM Klenow fragment (KF) were added to an eppendorf tube. Reactions were conducted for one hour at room temperature and stopped by adding 0.1 M EDTA. Data of FRET efficiency (E^*) are graphically shown on histograms, which describe distribution and changes of molecule species occupancy. Fig. 3.25 shows E^* before (top panel) and after polymerization (bottom panel). Table 13 summarizes the data and introduces the subsequent difference of FRET efficiency (dE) between the initial and the final value.
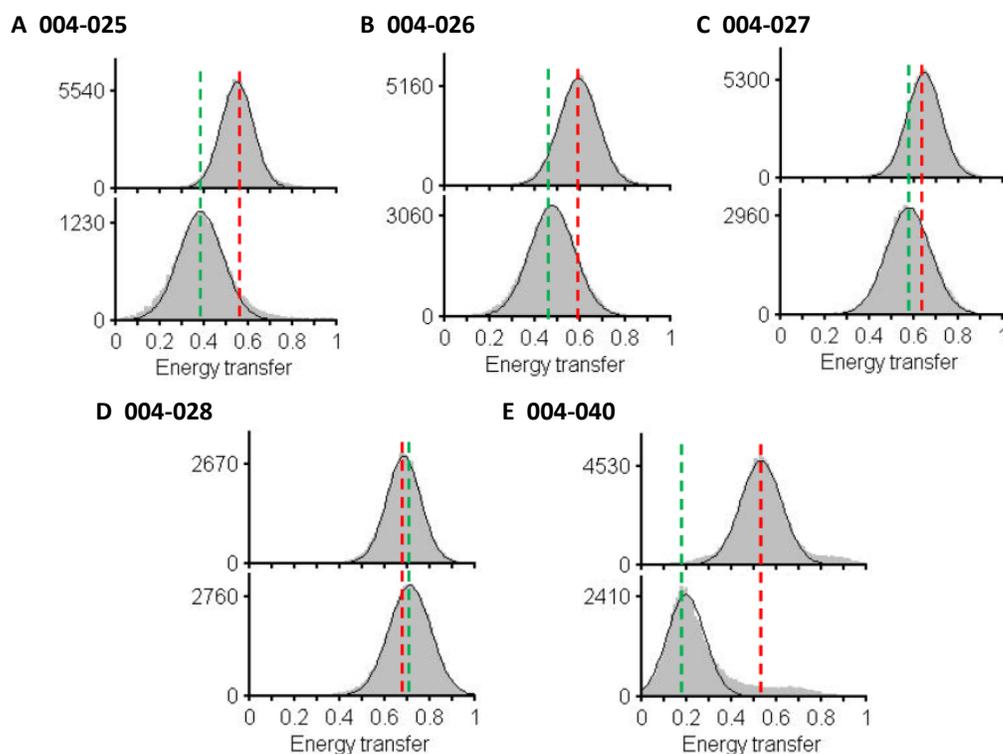
Figure 3.25: Stackplots of FRET efficiency (E^*) before (top) and after (bottom) polymerization of 10 nM 004-025 (**A**), 004-026 (**B**), 004-027 (**C**), 004-028 (**D**) and 004-040 (**E**) DNA constructs for one hour at room temperature and using for 40 nM KF and 100 µM of each dNTP (dATP, dTTP, dGTP and dCTP). Red and green dotted-lines represent E^* before and after one hour of polymerization, respectively.

DNA	004-025 (A)	004-026 (B)	004-027 (C)	004-028 (D)	004-040 (E)
E^*_0	0.55	0.6	0.65	0.69	0.53
E^*_F	0.39	0.48	0.58	0.71	0.19
dE	-0.16	-0.12	-0.07	0.02	-0.34

Table 13: FRET efficiency *(E*)*, measured by smFRET/TIRF, and difference of FRET efficiency (d*E*)* values for 40 nM KF before (E^*_0) and after (E^*_F) polymerization of 10 nM (A) 004-025, (B) 004-026, (C) 004-027, (D) 004-028 and (E) 004-040 DNA substrates for one hour at room temperature using 100 µM of each dNTP (dATP, dTTP, dGTP and dCTP).

The endpoint assay was repeated using the 004-025 DNA substrate for various polymerases: 40 nM Klenow fragment, 40 nM HoLaMa, 40 nM α subunit or 40 nM αεθ. Endpoint assays of this second experience were conducted under the same conditions than the precedent protocol (2.2.27). Fig. 3.26 represents graphically the values of *E** before and after polymerization. Data of *E** and d*E* are summarized in Table 14.
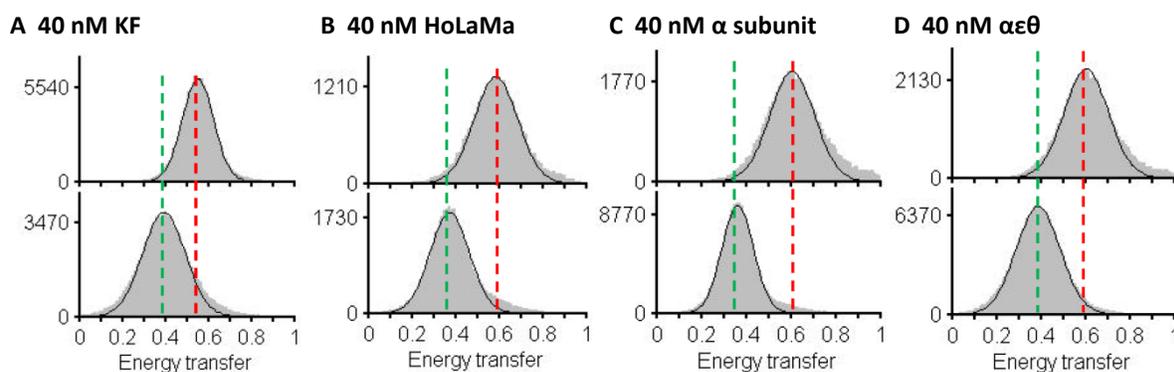
Figure 3.26: Stackplots of FRET efficiency (*E**) before (top panel) and after (bottom panel) polymerization of 10 nM 004-025 DNA construct using for KF (**A**), HoLaMa (**B**), α subunit (**C**) and αεθ (**D**) and 100 µM of each dNTP (dATP, dTTP, dGTP and dCTP)for one hour at room temperature. Red and green dotted-lines represent *E** before and after one hour of polymerization, respectively.

004-025	KF (A)	HoLaMa (B)	α subunit (C)	αεθ (D)
E^*_0	0.58	0.58	0.61	0.6
E^*_F	0.38	0.37	0.38	0.38
dE	-0.20	-0.21	-0.23	0.22

Table 14: FRET efficiency *(E*)* and difference of FRET efficiency (d*E)* values for KF(A), HoLaMa (B), α subunit (C) and αεθ (D) before (E^*_0) and after (E^*_F) polymerization of 10 nM 004-025 DNA substrate for one hour at room temperature and in the presence of 100 µM of each dNTP (dATP, dTTP, dGTP and dCTP).

3.7.2 Binding experiments

Binding experiments were carried out according to the procedure described in paragraph 2.2.32. In addition to histograms, time traces, or fluorescence intensity trajectories, represent the change in FRET efficiency for each single molecule. Trajectories allow analyzing some events such as protein binding or DNA polymerization, complementing histograms to understand the behavior of each assayed polymerase. A stable high E^* value corresponds to a continuous protein attachment to the DNA, whereas a peak represents a binding event. The assay was very reproducible if we consider that the initial value of E^* (E^*_0) was always about 0.54 (DNA 004-025) and 0.56 (DNA 00-040) after DNA immobilization, as shown in Tables 15 and 16 (more binding traces in Appendix III).

3.7.2.1 004-025 DNA substrate

Table 15 shows the change of E^* as a consequence of 2 nM polymerase (KF, HoLaMa, α subunit or αεθ) binding to the 004-025 DNA substrate. Data from Table 15 is graphically represented in Fig. 3.27.1 and 3.27.2.
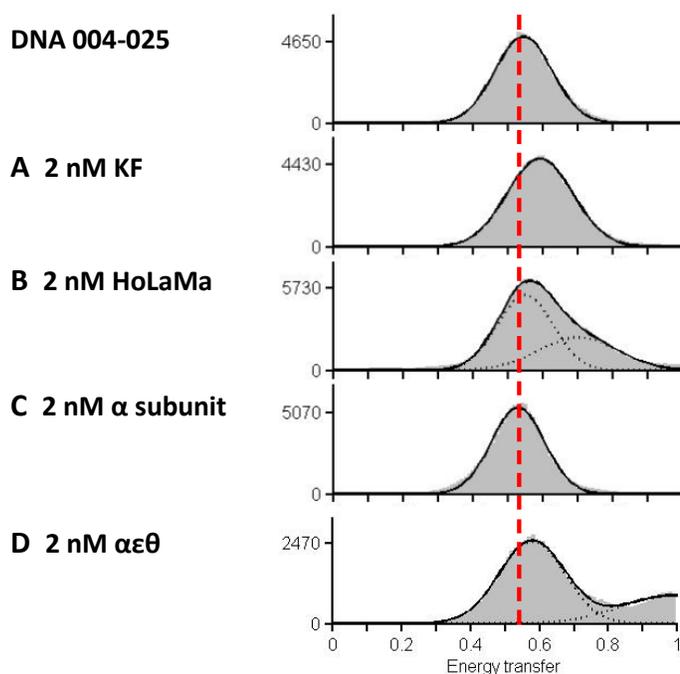
Figure 3.27.1: Stackplots of FRET efficiency (E^*) after KF (**A**), HoLaMa (**B**), α subunit (**C**) and αεθ (**D**) binding to 004-025 DNA construct. Red dotted-line represents binding sensor E^*. Top panel represents the immobilized DNA 004-025 molecules in the absence of any polymerase.

	004-025	KF (A)	HoLaMa (B)	α subunit (C)	αεθ (D)
E^*_0	0.54	0.54	0.55	0.53	0.55
E^*_B	-	0.59	0.57 / 0.71	0.53	0.58 / 0.96

Table 15: FRET efficiency *(E*)* values for 004-025 DNA substrate before (E^*_0) and after (E^*_B) addition of KF, HoLaMa, α subunit and αεθ binding.

A 2 nM KF

B 2 nM HoLaMa

C 2 nM α subunit

D 2 nM αεθ
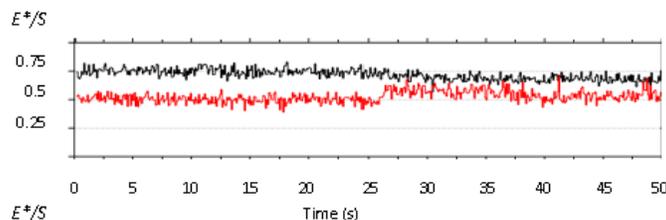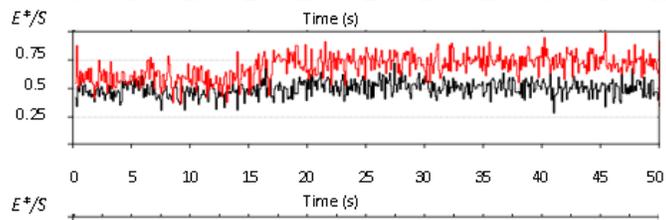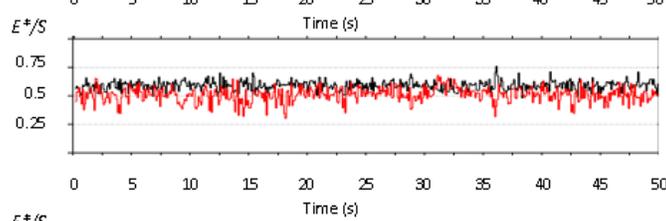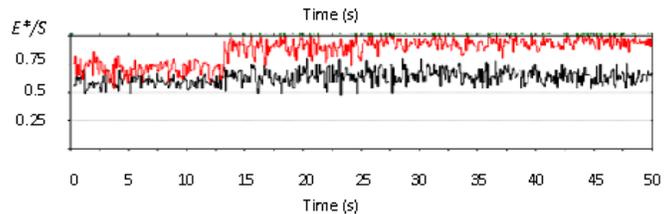
Figure 3.27.2: Exemplary FRET time traces representing binding of DNA polymerases to the 004-025 DNA construct: **(A)** 2 nM KF. Binding for about 27s leads to a small change in *E**; **(B)** 2 nM HoLaMa. Binding for about 33s leads to a small increase of *E**; **(C)** 2 nM α subunit. Binding leads to small change in *E*;* **(D)** 2 nM αεθ. Binding for about 36s leads to a large change in *E**. Black line represents stoichiometry; red line represents *E**.

When the 004-025 construct was used as binding sensor, the initial *E** value was about 0.54. However, the amount of shifting of *E** was different for all four assayed polymerases. Klenow fragment showed a shift of *E** after protein binding (from 0.54 to 0.59) (Fig. 3.27.A). Moreover, an increase in FRET efficiency and a considerable stable binding was distinguished in the time traces of KF, (Fig. 3.27.A; Appendix III). HoLaMa binding showed an additional shoulder of FRET efficiency

(0.71) (Fig. 3.27.B), with time traces in which binding was not as stable as in the case of KF because of the shorter periods of binding.

In the case of αεθ, the shoulder that appear after protein binding was much more remarkable (E^*=0.96) (Fig. 3.27.D) than that of HoLaMa. αεθ time traces showed an increase in FRET efficiency close to 1 when protein bound the 004-025 construct. Protein binding was very stable and with a shift towards 0.98 (Fig. 3.27.D; Appendix III). In contrast, α did not present a significant difference in FRET efficiency after protein binding with the same 004-025 DNA construct (Fig. 3.27.C), neither on the histogram nor on the time traces. Short trajectories of low E^* were perceived on α subunit time traces (Fig. 3.27.C; Appendix III).

3.7.2.2 004-040 DNA substrate

Table 16 summarizes E^* before (E^*_0) and after (E^*_F) binding of 2 nM polymerase (KF, HoLaMa, α subunit or αεθ) to the 004-040 DNA substrate. Histograms and time traces (Fig. 3.28) represent the data from Table 16.
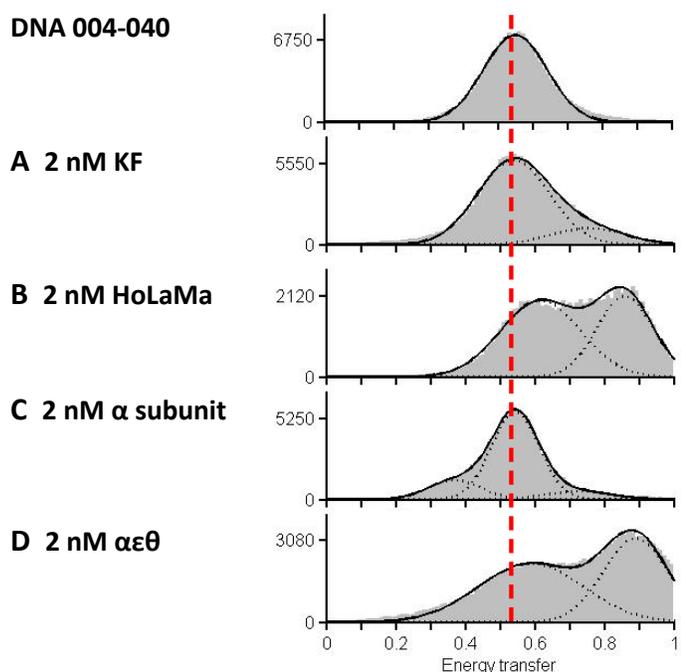
Figure 3.28.1: Stackplots of FRET efficiency (E^*) after the binding of KF (**A**), HoLaMa (**B**), α subunit (**C**) and αεθ (**D**) to the 004-040 DNA construct. Red dotted-line represents binding sensor E^*. Top panel represents the immobilized DNA 004-040 molecules in the absence of any polymerase.

	004-040	KF (A)	HoLaMa (B)	α subunit (C)	αεθ (D)
E^*_0	0.55	0.55	0.57	0.55	0.56
E^*_B	-	0.55 / 0.78	0.57 / 0.84	0.39 / 0.55 / 0.72	0.56 / 0.88

Table 16: FRET efficiency *(E*)* values for 004-040 DNA substrate before *(E^*_0)* and after *(E^*_B)* addition of KF, HoLaMa, α subunit and αεθ.
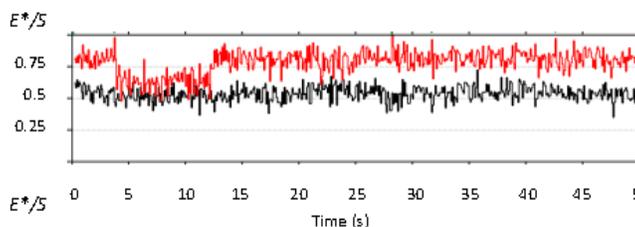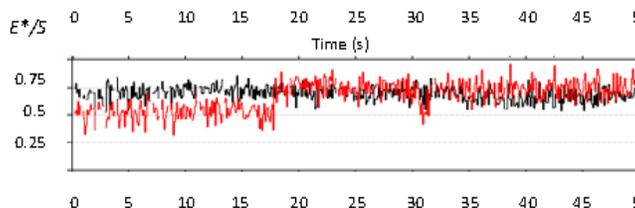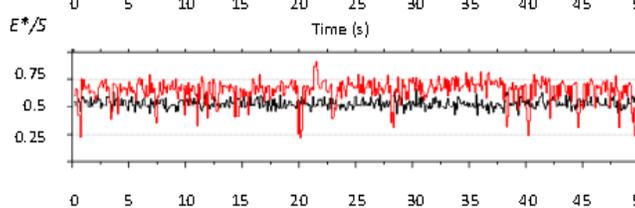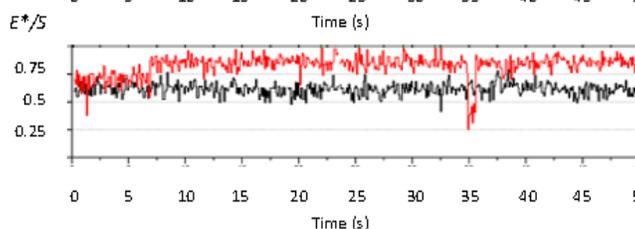
Figure 3.28.2: Exemplary FRET time traces representing binding of DNA polymerases to the 004-040 DNA construct: **(A)** 2 nM KF. Binding for about 4s followed by protein dissociation and a re-binding for another 37s; **(B)** 2 nM HoLaMa. Binding for 22s leads to a significant change in *E**; **(C)** 2 nM α subunit, changes in *E** represent binding and dissociation and three FRET levels (low, middle and high) are detected; **(D)** 2 nM αεθ. Binding for 30s, dissociation (1 s) and new binding for another 18s. Black line represents stoichiometry; red line represents *E**.

Interestingly, each of the four DNA polymerases led to different histogram of transfer efficiencies and at least a shoulder appeared after protein binding (Fig. 3.28). *E** of the binding sensor without any polymerase in solution was about 0.55-0.57. KF binding showed a shoulder with a peak position of *E** at 0.78 (Fig. 3.28.1.A). According to time traces, binding was very stable, although in

some traces protein dissociation was also present, as shown in Fig. 3.28.2.A. The shoulder observed for HoLaMa was with an E^* value equal to 0.84. Like in the case of binding to 004-025 DNA, protein binding was not as stable as KF.

Also αεθ showed a shoulder of E^* at 0.88. It is worth highlighting that a significant shift of E^*, from 0.56 to 0.88, was seen on time traces after protein binding and traces showed a stable binding. In contrast, two new unexpected peaks of E^*, one higher ($E^*=0.72$) and one lower ($E^*=0.39$) than the original one (0.55), appeared after the binding of α subunit to the DNA sensor 004-040 (Fig. 3.28.C). Accordingly, different high and low E^* frames appeared on the time traces.

3.7.2.3 003-025-032 DNA substrate

One nucleotide gapped-DNA was used as binding sensor for Klenow fragment and HoLaMa titration. Increasing concentrations of these two proteins were used with the aim of determining dissociation constant K_D. E^* of the 003-025-032 binding sensor was 0.4 whereas E^* raised after protein binding close to 0.6, in the presence and in the absence of complementary dTTP (Table 17; Fig. 3.29.1). For HoLaMa, the occupancy of molecule species bound to the DNA increased from 27.4 to 81% when the complementary dTTP was present in the imaging buffer (Fig. 3.29.1 bottom panel), improving also the stability of the binding (Fig. 3.29.2).
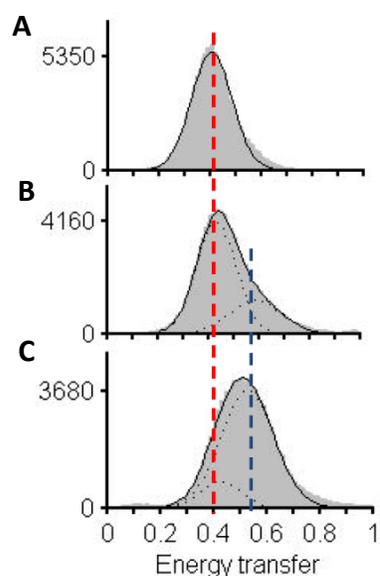
004-025	KF
E^*_0	0.4
E^*_B	0.58
$E^*_{B\_dTTP}$	0.57

Table 17: FRET efficiency *(E*)* values for 004-040 DNA substrate before (E^*_0) and after binding of 2 nM HoLaMa in the absence (E^*_B) and presence ($E^*_{B\_dTTP}$) of 100 µM dTTP.

Figure 3.29.1: Stackplots representing E^* of 003-025-032 binding sensor (**A**) and 2 nM HoLaMa binding in the absence (**B**) or presence (**C**) of 100 µM complementary dTTP. Red and blue dotted-lines represent E^* DNA sensor ($E^*=0.4$) and protein binding ($E^*=0.6$), respectively.

A 2 nM HoLaMa   *E*/S*

B 2 nM HoLaMa + 100 µM dTTP
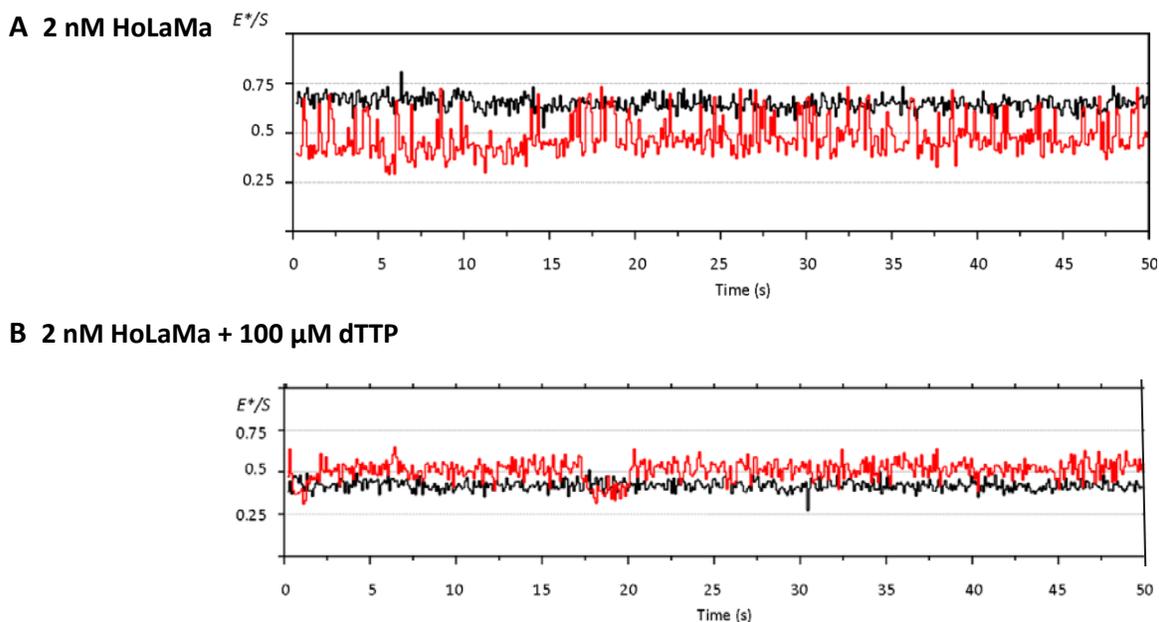
Figure 3.29.2: Exemplary FRET time traces representing binding of HoLaMa to the 003-025-032 DNA construct, in the absence **(A)** or in the presence **(B)** of 100 µM dTTP, complementary to the nucleotide of the gap. In the absence of dTTP, HoLaMa bound the DNA only for some seconds, represented by peaks of *E**; on the contrary, the presence of dTTP leads to the traces of 16-30s because of the stable binding to the DNA. Black line represents stoichiometry; red line represents *E**.

3.7.3 Real time polymerization assays

The evaluation of the polymerase activity of Klenow fragment, HoLaMa, α subunit and αεθ was performed in real time, as previously described in paragraph 2.2.33. The objective of the real time evaluation was to determine changes of occupancy of different FRET species as a consequence of DNA polymerization, using those DNA substrates that presented the most significant change from the initial state to the final state after polymerization in the endpoint assay. This new experiment was carried out in real time meaning that non-polymerized DNA molecules were immobilized on the glass surface and imaged before the polymerases were added. From the endpoint assays we expected values before and upon polymerization of about *E** equal to 0.54 and 0.35 for 004-025 DNA substrate, and 0.55 and 0.2 for 004-040 DNA substrate. Different concentrations of polymerases were used for obtaining an observable changing signal, taken into account that the each polymerization can only be monitored once. Additional polymerization time-traces are found in Appendix IV.

3.7.3.1 004-025 DNA substrate

Table 18 presents results of E^* before (E^*_0) and after (E^*_P) the real time polymerization of 20 pM 004-025 DNA substrate, carried out by 4 different polymerases and using 100 µM dNTPs in the imaging buffer. Data are graphically shown in Fig. 3.30.1, in which peaks represent E^* before (Fig. 3.30.1 top panel) and after three minutes (Fig. 3.30.1 middle panel) and ten minutes (Fig. 3.30.1 bottom panel) of polymerization for all four polymerases. Time traces of polymerization events for 004-025 DNA substrate are shown in Fig. 3.30.2. Polymerase concentration within the imaging buffer was 500 pM for KF, 2 nM for HoLaMa, 50 pM for α subunit and 50 pM for αεθ.
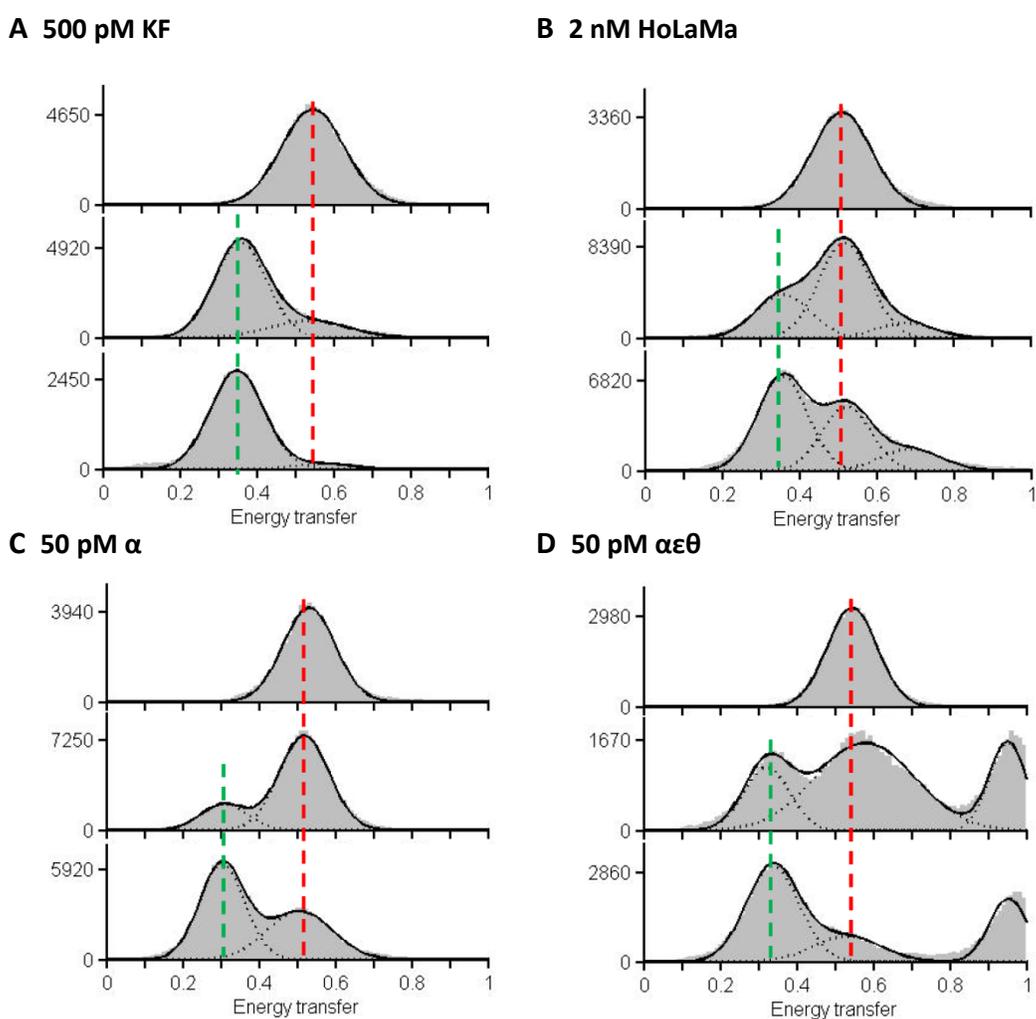
Figure 3.30.1: Stackplots of FRET efficiency (E^*) of 004-025 construct (top panel) and after three minutes (middle panel) and ten minutes (bottom panel) of polymerization in the presence of 100 µM each dNTP. Assayed polymerases were 500 pM KF (**A**), 2 nM HoLaMa (**B**), 50 pM α subunit (**C**) and 50 pM αεθ (**D**). Red and green dotted-lines represent E^* of 004-025 DNA construct and polymerization, respectively.

004-025	KF (A)	HoLaMa (B)	α subunit (C)	αεθ (D)
E^*_0	0.54	0.52	0.53	0.55
E^*_P	0.35 / 0.54	0.32 / 0.52 / 0.7	0.31 / 0.53	0.33 /0.57 / 0.95

Table 18: FRET efficiency *(E*)* values for 004-025 DNA substrate before (E^*_0) and after 10 minutes of polymerization (E^*_P) by 500 pM KF, 2 nM HoLaMa, 50 pM α subunit and 50 pM αεθ in real real time, using 100 μM dNTPs.

Overall, all four polymerases showed a shift of *E** to the left, from 0.54 to 0.35 and the diverse peaks present in Fig. 3.30.1 determine the evolution of molecules species occupancy from a not-polymerized state (high-FRET level) to a full-polymerized state (low-FRET level, equal to 0.35). The shift towards 0.35 was expected for polymerized DNA molecules, according to the endpoint assay (section 3.7.1). In the case of HoLaMa and αεθ, another peak of *E** was present equivalent to 0.7 (Fig. 3.30.1B) and 0.95 (Fig. 3.30.1D), respectively, corresponding to different states of incomplete polymerization or intermediate conformations as explained in the following chapter (section 4.4).

Klenow fragment showed a single peak of FRET efficiency equal to 0.35 after ten minutes in real time (Fig. 3.30.1 A), indicating that practically all molecules were fully polymerized. Time traces of Klenow fragment (Fig. 3.30.2.A) showed a slight increase of *E** upon protein binding, immediately followed by a pronounced shift that represented the polymerization reaction, up to the above-mentioned value of *E** equal to 0.35. For HoLaMa, the relative number of full-polymerized molecules (low-FRET level) after 10 minutes of polymerization was significantly lower than for Klenow fragment. Moreover, HoLaMa time traces were characterized by a very slow reaction (Fig. 3.30.2.B), representing the low speed of polymerization of the enzyme. In some of them, the shift of *E** to the lower value took almost 13-30s (Fig. 3.30.2.B; Appendix IV) and some short increases in *E** were observed in the range of the fall, presumably representing protein dissociation and new binding.
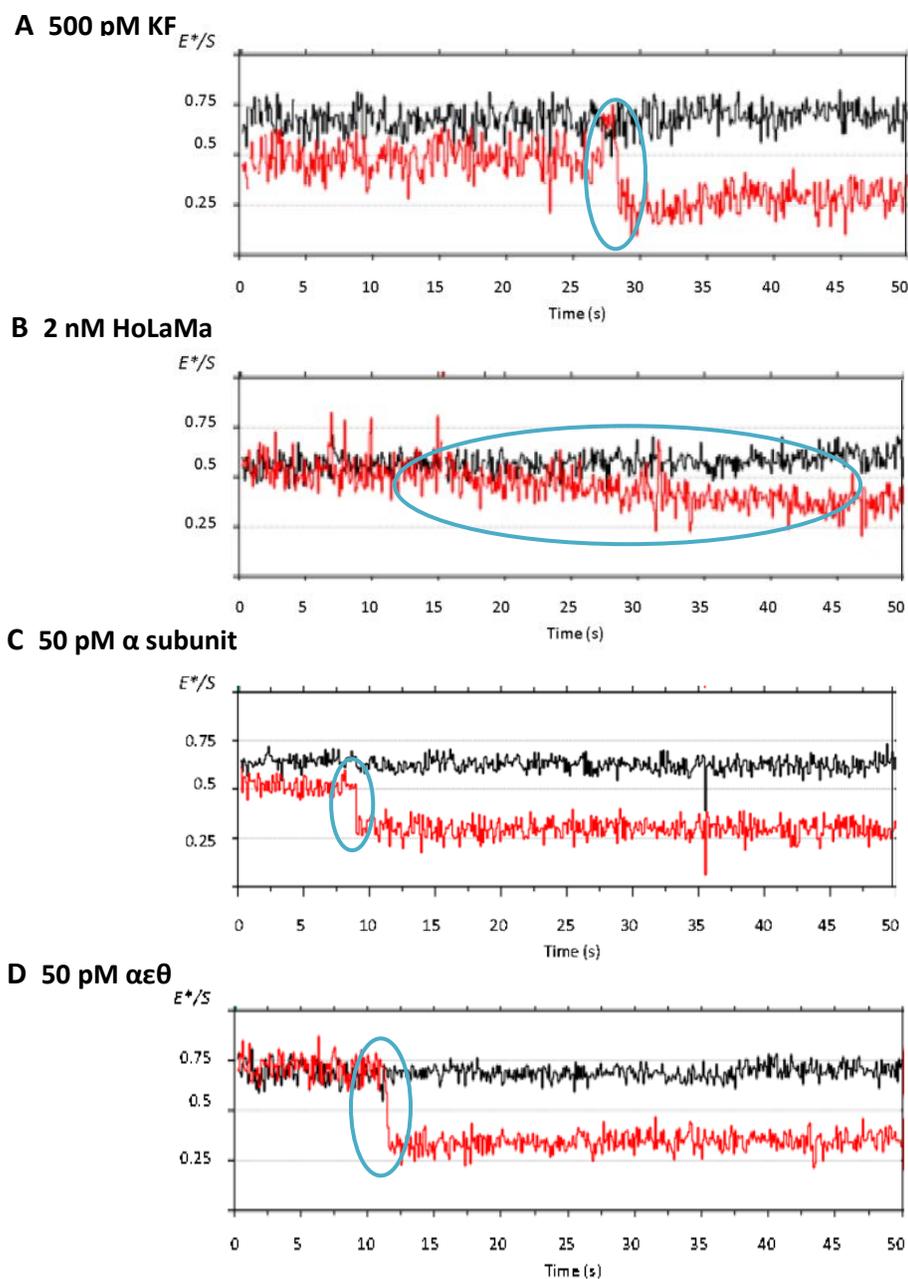
Figure 3.30.2: Exemplary FRET time traces representing polymerization in real time of the 004-025 DNA construct in the presence of 100 μM of each dNTP. **(A)** 500 pM KF, binding for around 2-3s with a short increase of FRET efficiency followed by a decrease to 0.35 representing polymerization; **(B)** 2 nM HoLaMa, shift of E^* from 0.55 to 0.35 in about 27s; **(C)** 50 pM α subunit, significant decrease of E^* from 0.55 to 0.35 in approximately 2s; **(D)** 50 nM αεθ, decrease of E^* from 0.75 to 0.35 in less than 2s. Black line represents stoichiometry; red line represents E^*. The blue ellipses highlight the polymerization events.

In the presence of 50 pM of α subunit, direct shifts from E^*=0.53 to 0.31 were observed, indicating polymerization. On time-traces, the change in E^* was even faster (about 1s) than KF and did not present an increase of E^* before polymerization (Fig. 3.30.2.C), as observed on time-traces of KF. A change in the occupancies from 0.95 to 0.33 was visualized in the case of αεθ (Fig. 3.30.1.D). The decrease to the expected value of E^* (0.35) was even faster (about 0.5 s) than that of α subunit (Fig. 3.30.2.D).

3.7.3.2 004-040 DNA substrate

The real time polymerization E^* values for the 004-040 are summarized in Table 19 and graphically represented in Fig. 3.31.1 and 3.31.2. Histogram represents DNA E^* before (Fig. 3.31.1 top panel) and after three minutes (Fig. 3.31.1 middle panel) and ten minutes (Fig. 3.31.1 bottom panel) of polymerization.

The polymerization activity of various DNA polymerases (50 pM KF, 2 nM HoLaMa, 50 pM α subunit and 50 pM αεθ), was monitored (Fig. 3.31.1). A peak of E^* equal to 0.2 indicated the increasing population of full polymerized molecules. This 0.2 value is in perfect agreement with the one obtained in the corresponding endpoint assay (paragraph 3.7.1). The shift comes from the original E^* at 0.56, related to those not polymerized DNA molecules. As present for the 004-025 DNA substrate, HoLaMa and αεθ featured a shoulder of E^* at 0.88 (Fig. 3.31.1B) and 0.9 (Fig. 3.31.1D), respectively.

Similar to the results obtained for 004-025 DNA substrate, Klenow fragment completed the polymerization of most molecules within 10 minutes as indicated by the large fraction of molecules showing a FRET efficiency around 0.2 (Fig. 3.31.1.A). Time traces of Klenow fragment (Fig. 3.31.2A) showed a peak of E^* corresponding to protein binding followed by a decrease up to the expected E^* value (0.2). This decrease was faster for concentration of dNTPs equal to 100 μM, and slower for lower concentrations of nucleotides (1 μM) (Fig. 3.31.3).
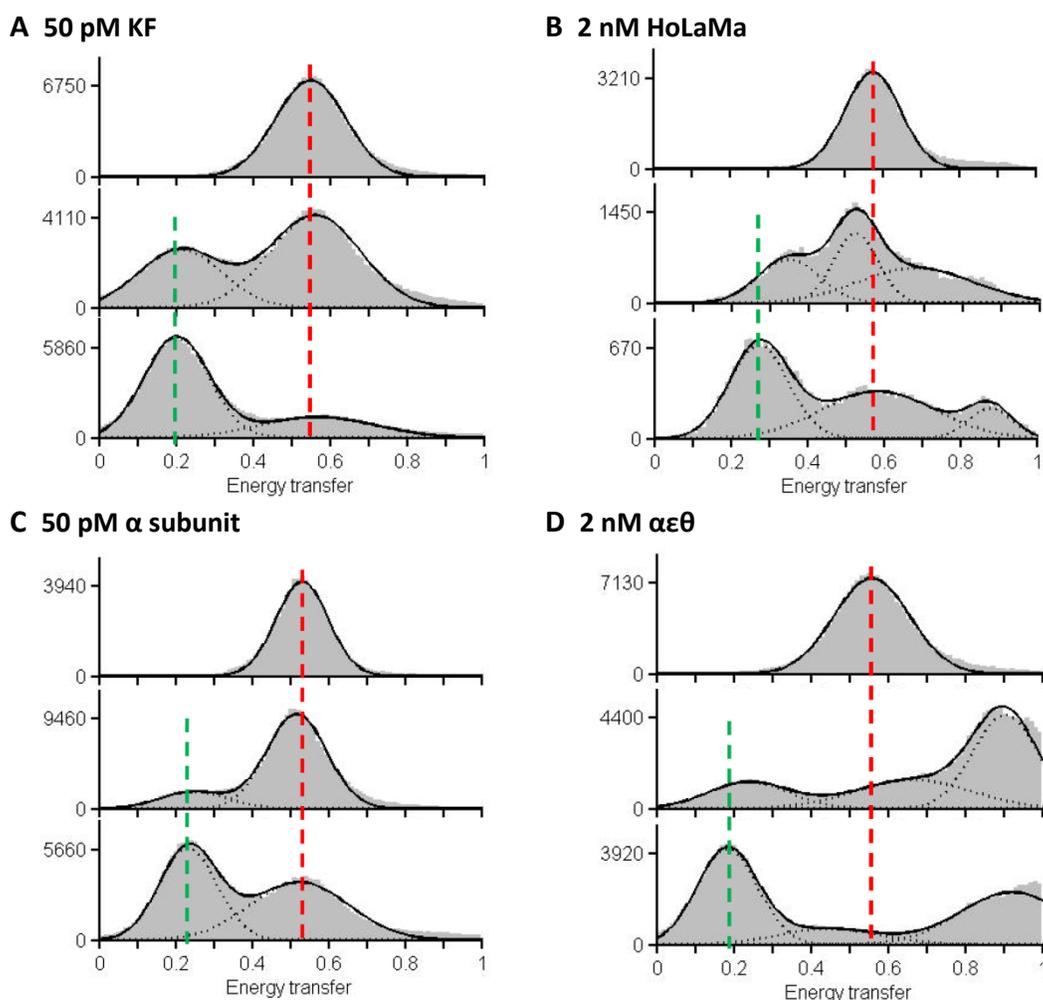
Figure 3.31.1: Stackplots of FRET efficiency (E^*) of 004-040 construct (top panel) and after three minutes (middle panel) and ten minutes (bottom panel) of polymerization with 100 µM dNTPs. Histograms show results after using 50 pM KF (**A**), 2 nM HoLaMa (**B**), 50 pM α subunit (**C**) and 50 pM αεθ (**D**). Red and green dotted-lines represent E^* of 004-040 DNA construct and polymerization, respectively.

004-040	KF (A)	HoLaMa (B)	α subunit (C)	αεθ (D)
E^*_0	0.55	0.57	0.55	0.56
E^*_P	0.2 /0.55	0.28 / 0.58 / 0.88	0.22 / 0.55	0.20 /0.56 / 0.9

Table 19: FRET efficiency (E^*) values for 004-040 DNA substrate before (E^*_0) and after 10 minutes of polymerization (E^*_P) by 50 pM KF, 2 nM HoLaMa, 50 pM α subunit and 50 pM αεθ in real real time, using 100 µM dNTPs.
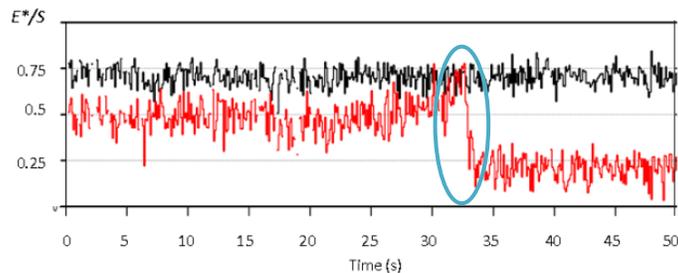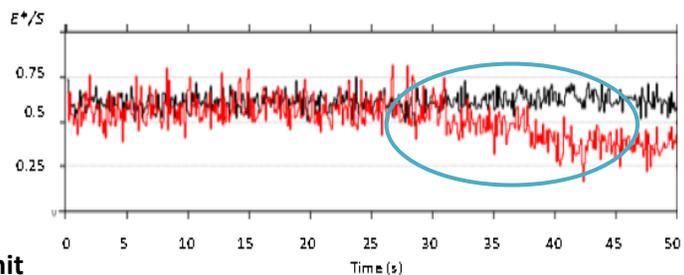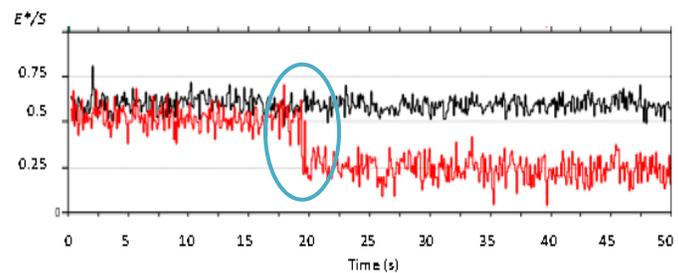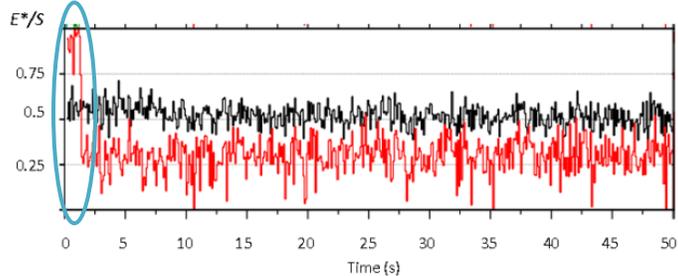
Figure 3.31.2: Exemplary FRET time traces representing polymerization of the 004-040 DNA construct with 100 µM dNTPs: **(A)** 50 pM KF in solution, binding for around 2s followed by a change in FRET efficiency to 0.2 representing DNA polymerization, **(B)** 2 nM HoLaMa, shift of E^* from 0.55 to almost 0.2 in about 16s; **(C)** 50 pM α subunit, significant decrease of E^* from 0.55 to 0.2 in 2s, approximately; **(D)** 50 nM αεθ, severe decrease of E^* from 0.9 to 0.2 in less than 2s. Black line represents stoichiometry; red line represents E^*. The blue ellipses highlight the polymerization events.

On the contrary, molecule populations regarding medium-FRET (0.58) and higher-FRET (0.88) levels were considerably larger after 10 minutes of polymerization by HoLaMa, in comparison to Klenow fragment (Fig. 3.31.1.B). Accordingly, HoLaMa time traces showed again a very slow shift of E^* from the upper level to the lower one (Fig. 3.31.2.B), indicating the low speed of reaction of this artificial polymerase.

α subunit featured a shift of E^* from 0.55 to 0.22 for full-polymerized molecules occupancy (Fig. 3.31.2.C). As for 004-025 DNA substrate, time traces showed that the shift of E^* was fast (about 1s) and without a previous increase of E^* regarding polymerase binding (Fig. 3.31.2.C), as observed on Klenow fragment time traces (Fig. 3.31.2.A). On the other hand, αεθ showed three populations with corresponding peaks around E^* at 0.22, 0.55 and 0.89 (Fig. 3.31.1.D). The peak at the lowest level of FRET efficiency characterized those molecule species in which polymerization is complete whereas the other two peaks represent not-polymerized molecules. The presence of a third peak at 0.90, which was not observed for the α subunit alone, could be related to an incomplete polymerization and/or intermediate conformations. On time traces, the shift of E^* up to 0.2 was achieved in one or two steps (Fig. 3.31.2.D). Each polymerization was completed in less than 1s.
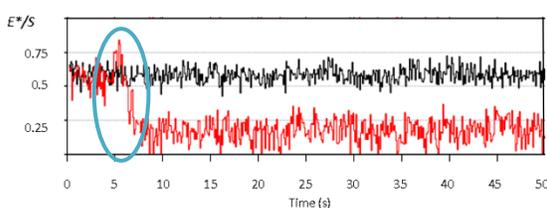
A 50 pM KF + 100 µM each dNTP   **B 50 pM KF + 1 µM each dNTP**
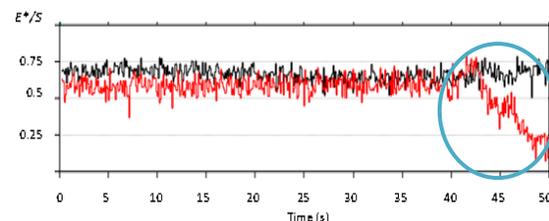
Figure 3.31.3: Exemplary FRET time traces representing polymerization of the 004-040 DNA construct with 50 pM KF and two different concentrations of dNTPs: **(A)** 100 µM dNTPs, peak of binding immediately followed by a change in FRET efficiency to 0.2 representing DNA polymerization that takes part in about 1s; **(B)** 1 µM dNTPs, slight increase of E^* followed by a decrease of FRET efficiency up to 0.2 in two events, separated by a peak of E^* representing new binding event after protein dissociation. The decrease of E^* was for 3s each. Black line represents stoichiometry; red line represents E^*. The blue ellipses highlight the polymerization events.
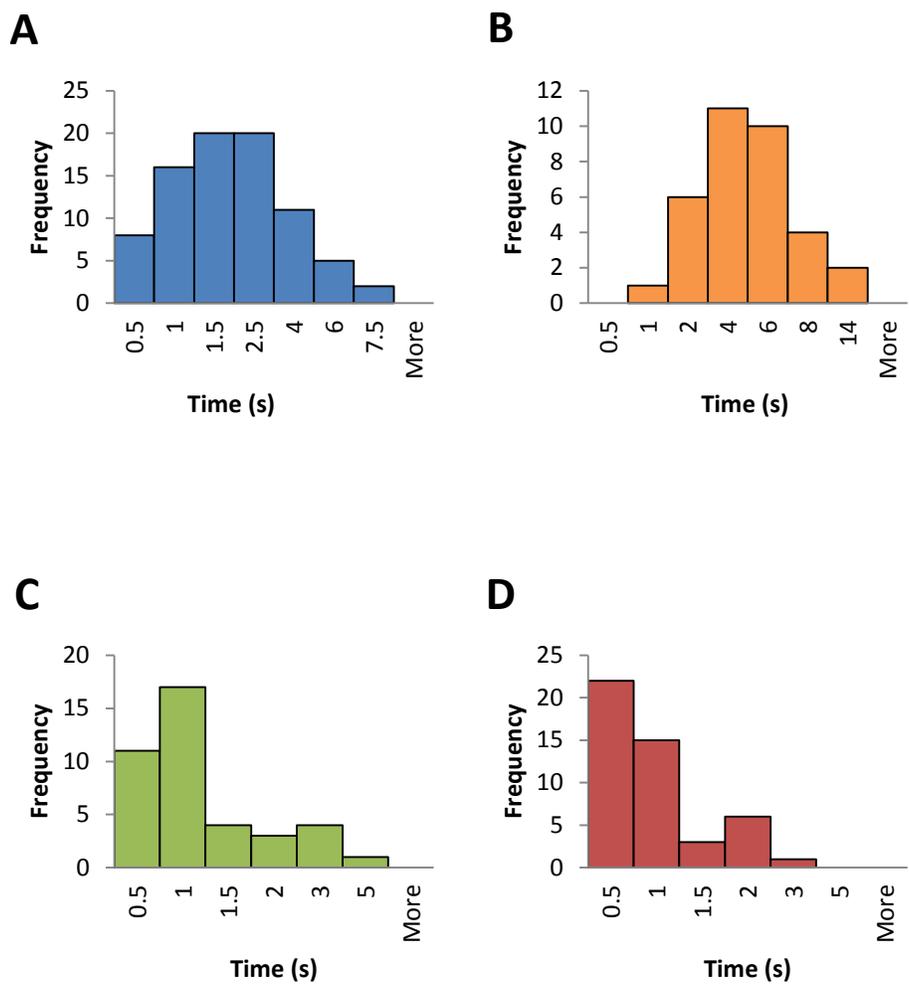
Figure 3.32: Histograms representing the frequency of duration of each polymerization event for all four polymerases in the presence of 20 pM DNA and 100 µM dNTPs. (A) 50pM KF, (B) 2nM HoLaMa, (C) 50 pM α subunit and (D) 50 pM αεθ.

Figure 3.32 represents the average duration for all polymerization events observed on time traces for the different assayed polymerases. The average duration was 2 seconds for 50 pM of Klenow fragment, 5 seconds for 2 nM of HoLaMa, 1 second in the case of 50 pM of α subunit and 0.5 seconds for 50 pM of αεθ. Therefore, the average nucleotide incorporation rate for each polymerase was 12 nucleotides per second in the case of Klenow fragment, 5 nucleotides per second for HoLaMa, 25 nucleotides per second for α subunit and 50 nucleotides per second in the case of αεθ, considering the presence of 25 extending nucleotides in the ssDNA.

CHAPTER 4 – DISCUSSION

4.1 Trimeric sub-assembly of E. coli DNA Pol III

DNA Pol III holoenzyme is the only essential DNA polymerase in *E. coli* (Maki & Kornberg, 1988). This holoenzyme is composed of ten types of subunits assembled in three different complexes: the catalytic core, the loading clamp and the clamp loader (McHenry, 2011). However, the complexity of DNA Pol III and the absence of crystal structures make highly difficult the study and structural considerations of the whole DNA Pol III holoenzyme and even some of its individual subunits. That is the case of α subunit, a partial structure of which (Lamers *et al.,* 2006), lacking the last 243 residues, showed the classic right hand folding, typical for all DNA polymerases, consisting of the Palm, Thumb and Fingers domains. Moreover, a cluster of arginines, partially exhibited at the Palm (R390 and R396) and in part at the Fingers (R709 and R710) domains, is considered to interact with the γ phosphate of the incoming nucleotide. On the other hand, the N-terminus represents the PHP domain, characterized by an α/β distorted barrel structure, whilst the C-terminal region is still not structurally resolved for the DNA polymerase III α subunit of *E. coli*.

In the case of the αεθ, structural data are based on incomplete information from partial structures of individual subunits, all of which are proved to be particularly difficult to isolate and purify. In the literature, several protocols have been developed for individual subunit purification, both from a crude protein extract (McHenry & Crow, 1979) or from separately-purified individual subunits and then assembled *in vitro* (Studwell-Vaughan & O'Donnell, 1993). A simultaneous protein overexpression of the three proteins of the αεθ complex in a single host was then reached by cloning the three coding genes, *dnaQ* (ε), *holE* (θ) and *dnaE* (α), into one vector under the control of the same promoter (Kim & McHenry, 1996). However, sequence cloning determined the effects of a transcriptional polarity that affects especially against the *dnaE* gene, considering its length of more than 3,400 base pairs (Tomasiewicz & McHenry, 1987) and its distance from the promoter. The resulting effect was thus the production of a small amount of α subunit and the accumulation of soluble ε-θ heterodimeric forms and ε-ε insoluble complexes, with the subsequent low αεθ yield. Owing to these reasons, and so as to improve the amount of purified protein, Conte (2012) performed a set of co-expression systems based on cloning the *dnaE* gene into the pBAD plasmid while the *dnaQ* and *holE* genes were cloned into the pGOOD plasmid. The cloning of the individual α subunit into an exclusive plasmid with a higher number of copies, compared to that sustaining the expression of ε and θ subunits, may compensate for the higher difficulty of its synthesis. The different induction system for the two vectors (arabinose and IPTG for pBAD and pGOOD, respectively) provided the ability to adjust the

production of all subunits independently. Taking advantage of these features, also the gene coding the τ subunit, *dnaX*, was cloned into the pGOOD vector, at the first position near the promoter to improve the yields of expression and purification of the ατεθ complex. Shine-Dalgarno sequence compensated ε subunit production.

Previously, *in vitro* studies had shown the preferential assembly of a clamp loader with three τ subunits that would allow the binding of three αεθ that, working simultaneously, would enhance considerably the efficiency of DNA Pol III activity (McInerney *et al.,* 2007) (Fig. 1.28). Some authors have determined the ability of a trimeric sub-assembly of DNA Pol III to improve the efficiency of DNA polymerization (McInerney *et al.,* 2007; Georgescu *et al.,* 2014) when compared to the classic dimeric model (McHenry, 1982); nevertheless, these trimeric complexes were reconstituted *in vitro* from unassembled subunits and there are not methods for a trimeric sub-assembly of the DNA Pol III *in vivo*.

In the present work, a procedure suitable for the overexpression and purification of a trimeric DNA Pol III, assembled *in vivo*, was developed by co-expressing α, τ, ε and θ subunits. Basically, after growing a pre-inoculum of one transformant colony in 1 ml of fresh LB, supplemented with appropriate antibiotics, at 37° C for 9 hours, a 1:500 dilution was performed and grown over-night under constant shaking (180 rpm) at 30° C. Protein expression was triggered with both arabinose and IPTG (1 mM each) at 30° C for 2.5 hours. The purification of the ατεθ complex was performed according to the procedure described in paragraph 2.2.10.

The trimeric sub-assembly of the DNA Pol III was identified by 4 tests. First of all, the three chromatograms of polymerase, ATPase and exonuclease activities, shown in Figures 3.12A, B and C, respectively, confirmed the presence of a peak of activity on fractions 13 and 14, corresponding to the elution volume of thyroglobulin (660 kDa), used as a standard to calibrate the gel filtration column. The similarity of the standard mass value to the expected molecular weight of a trimeric complex (about 710 kDa) makes consistent the idea that a trimeric sub-assembly of the DNA Pol III *in vivo* may spontaneously form. Therefore, this preliminary experiment represents the first experience in which the *in vivo* overexpression and trimeric sub-assembly of three αεθ and three τ subunits ($\alpha_3\tau_3\varepsilon_3\theta_3$) takes place.

Second, ESI mass spectrometry confirmed that the purified ατεθ complex presented a molecular mass equal to 708 kDa (Fig. 3.10), representing an *in vivo* trimeric assembly featuring three α subunits (129.9 kDa each), three τ subunits (71.1 kDa each), three ε subunits (27.5 kDa each) and two or three θ subunits (8.6 kDa each). The detected α-ε interactions confirmed previously described links between α N-terminus and ε

C-terminal regions (Taft-Benz & Schaaper, 1999), whereas the ε-θ links are related to the data regarding interactions between ε N-terminal region and θ subunit (De Rose *et al.,* 2003). The level of association of τ subunits, as shown by gel electrophoresis of the purified ατεθ complex (Fig. 3.8.A), was similar to other proteins present in the complex. Actually, the presence of three α subunits could be explained only by the presence of three τ subunits that, by binding between them, enable the three polymerases to associate to each other.
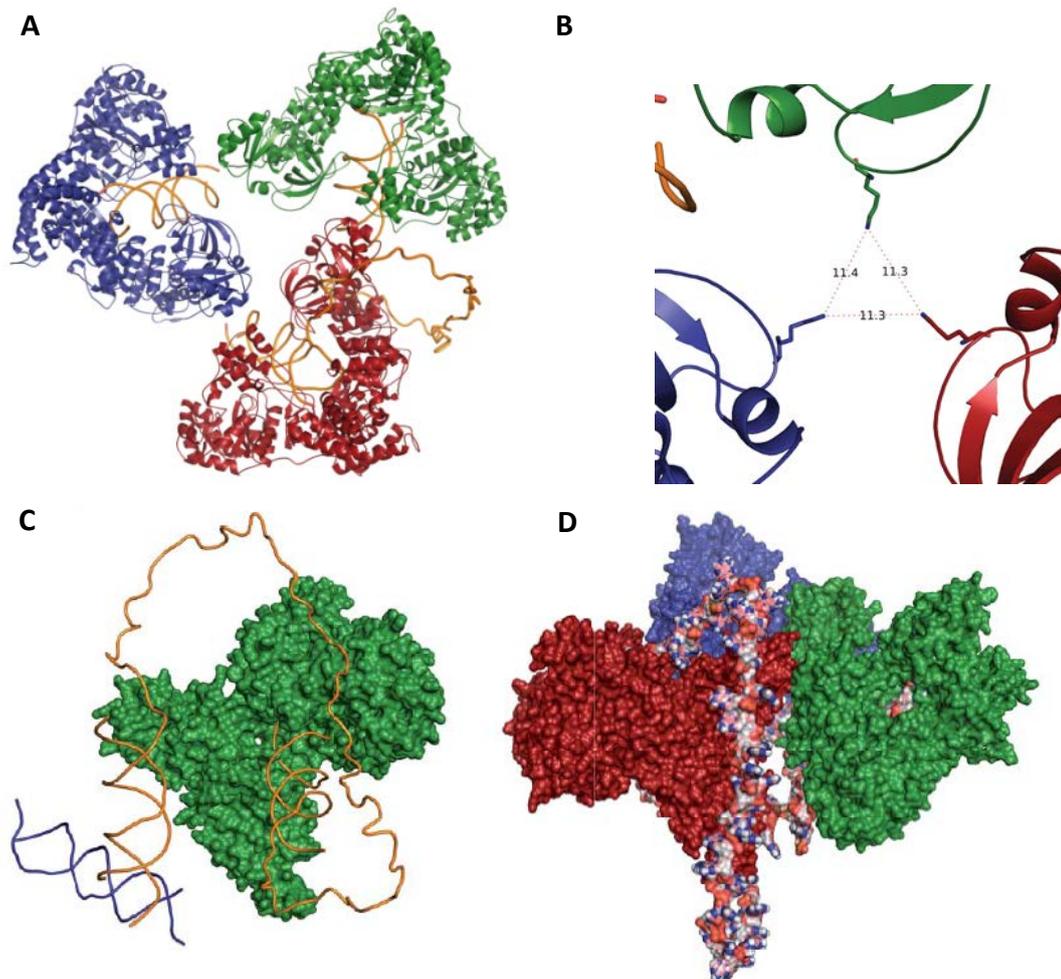
Figure 4.1: (A) Partial model of the trimeric replicase showing 3 α subunits associated with a single- and a double-primed DNA; **(B)** Detail of the molecular model of trimeric replicase observed in fig. 4.1.A. Each lysine and α subunit portion are reported with the same color used in fig. 4.1.A. We imposed a distance equal to 11.4 Å between the K1037 residue of each α subunit, since it is the length of the cross-linker (SB3); **(C)** Detail of the single- and the double-primed DNAs associated to a single α subunit, shown in green; **(D)** Side view of the molecular model reported in fig. 4.1.A. Structures were rendered with PyMol (The PyMOL Molecular Graphics System, version 1.3, Schrödinger, LLC). The tertiary structure of *E. coli* α subunit was obtained Swis-Model using as a template the *Thermus aquaticus* DnaE replicase (PDB 2hpm) (Arnold *et al.,* 2006). The dsDNA structures from *T. aquaticus* DnaE replicase (PDB 3e0d) were positioned superimposing the *E. coli* protein model and the *T. aquaticus* polymerase in complex with DNA (PDB 3e0d). The ssDNA (65 nucleotides) was generated with Foldit24.

Remarkably, the observed α-α interaction (engaging K1037 or Lys 1044; Table 9) is located in the C-terminus, which is absent in the partial crystal structure of α subunit (Lamers *et al.,* 2006). Taking into account these interactions between α subunits of the αεθ complex, a partial model of the trimeric replicase can be confirmed: only one α subunit is required for the elongation of the leading strand, whereas the replication of the lagging strand is sustained by two α subunits, whose simultaneous elongation activity depends on a sufficiently long single-stranded DNA connecting two active sites (Fig. 4.1).

Third, DNA polymerase activity was then assayed under steady-state conditions, using the trimeric replicase and the observed activity was compared to that of purified α subunit. A 40mer DNA template strand annealed to a single octameric primer was used as substrate, and dTTP was included as the single dNTP at a concentration equal to 100 μM. Under steady-state conditions, the three α subunits of the αεθ complex were found to exert their function simultaneously (Fig. 4.2.A). The k_{cat} was indeed equal to 1.89 ± 0.15 s^{-1} for α subunit and 2.59 ± 0.26 s^{-1} for the αεθ complex, suggesting a slight increase of polymerase activity in the trimeric complex, something expected when considering the stimulating effect of ε subunit on polymerase activity (Maki & Kornberg, 1987; Perrino *et al.,* 1999).

Similar observations were obtained when the 3' → 5' exonuclease activity of the trimeric replicase was tested with the *p*-nitrophenyl ester of thymidine 5'-monophosphate (pNP-TMP) (Hamdan *et al.,* 2002b). The three ε subunits of the αεθ complex performed their function simultaneously (Fig. 4.2.B). The k_{cat} was 9.66 ± 0.36 and 8.50 ± 0.29 s^{-1} for the catalytic αεθ and the trimeric replicase, respectively.
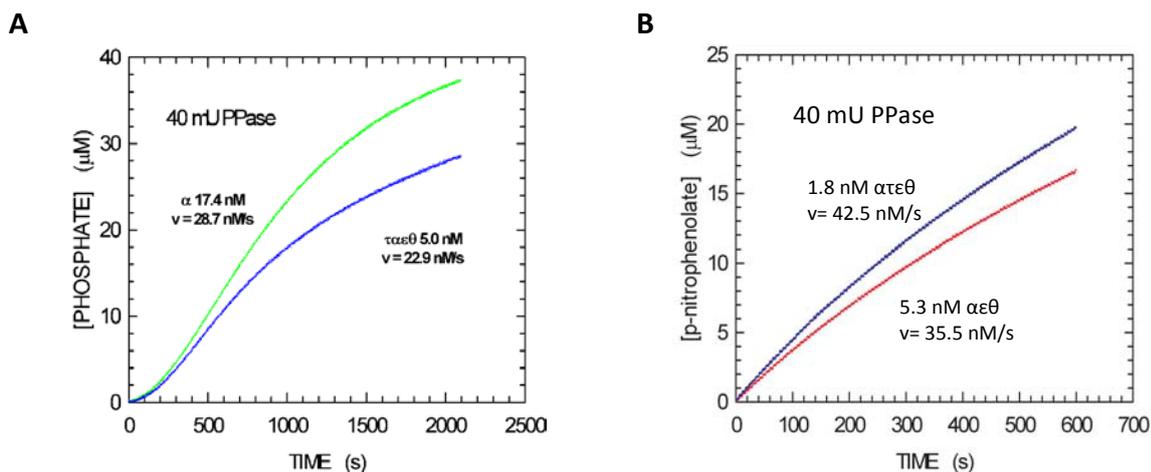
Figure 4.2: (A) Steady-state assay of DNA polymerase activity of 17.4 nM α subunit (green dots) and 3.2 nM αεθ complex (blue dots), in the presence of 1 μM 40mer DNA and 100 μM dTTP. **(B)** Steady-state assay of 3' → 5' exonuclease activity of 5.3 nM αεθ catalytic core (red dots) and 1.8 nM αεθ complex (blue dots), in the presence of 3.3 mM *p*-nitrophenyl ester of 5'-thymidine monophosphate (pNP-TMP).

Finally, a set of DNAs was used in single-turnover assays with the trimeric replicase in order to model the replication of the leading and lagging strands. The double-primed DNA was used as a model of lagging strand whereas the single-primed DNA mimicked the leading strand. Previously, under steady-state conditions it had been noticed that the highest DNA polymerase activity was observed, by far, in the presence of the 25mer DNA (13.6 nM/s) (Table 10). With regards to the evaluation of double-primed DNAs, the elongation of the 3' end, representing template head, was strongly inhibited and replicated slower than the 5' end (template tail) in the presence of the 50mer DNA, suggesting that a close downstream double-stranded DNA inhibits elongation. On the contrary, extension rates using the 100mer DNA substrate were much faster and no significant differences were found between the 100mer and 75mer DNA substrates (Table 10). Furthermore, the extension velocity of the 100mer DNA at the head site, using dTTP, was equal to 9 nM/s. When both dTTP and dATP were present, DNA polymerase activity increased up to 19 nM/s, indicating the concomitant extension of both head (3') and tail (5') of the template. Lastly, in the presence of dTTP, dATP, and dGTP, the extension rate increased, being equal to 38 nM/s (Fig. 3.20). This increase suggests that the activity related to the 25mer DNA elongation was equal to 19 nM/s, twice the activities observed for the double-primed DNA. Therefore, according to these observations, the activity of two α subunits on the double-primed DNA slowed down the extension of the single-primed DNA, since the elongation rate of α subunit in the presence of the 25mer DNA was 6 times higher than the rate determined in the presence of the 100mer DNA (Table 10; Fig. 3.18; Appendix II). Nevertheless, the trimeric αεθ complex was able to extend simultaneously three primers, asymmetrically annealed to two DNA templates.

The differences regarding elongation rates among three different double-primed DNAs are directly related to the inter-primer distance. Actually, single turn-over assays revealed significantly slower results for DNA elongation when 50mer or 75mer DNA substrates were used as substrates, in contrast with results obtained with a 100mer DNA. As a consequence, it can be concluded that a range between 40 and 65 nucleotides is the minimum distance value for the parallel elongation of two primers annealed to the same DNA template. These data correspond to about 260-420 Å, considering the average length of a single-stranded DNA (Murphy *et al.,* 2004). Yao *et al.,* (2000) determined that at least 10 nucleotides in the single-stranded DNA are necessary to make possible the binding of DNA Pol III holoenzyme to the template. Thus, compressing this single-stranded DNA to less than 40-65 nucleotides, but always over 10 nucleotides, the enzyme activity of α subunit is further inhibited for the extension of the lagging strand.

Overall, all these observations provide for the first time a kinetic and molecular model of a trimeric DNA replicase, *in vivo* assembled, which efficiently coordinates leading and lagging strand synthesis.

4.2 Characterization of the PHP domain of E. coli DNA Pol III

DNA replicases may contain additional domains, such as the Polymerase-Histidinol-Phosphatase (PHP) domain, which is often detected in those enzymes responsible for genome replication (Aravind & Koonin, 1998). In *E. coli*, the only DNA replicase containing a PHP domain is the α subunit of DNA Pol III, being this DNA Pol III the only essential polymerase for the replication of the entire genome of *E. coli* (Maki & Kornberg, 1988). Thus, the presence of the PHP domain in the only essential DNA polymerase hypothesizes a major role for this domain. Nevertheless, its specific activity has not been confirmed yet. What it is clear is that the PHP domain of the DNA Pol III of *E. coli* does not play an exonuclease activity since the exonuclease assay performed in the presence of the wild type α subunit did show lack of hydrolysis of the pNT-TMP (Fig. 3.21). Therefore, its activity is not comparable to that of the PHP domain of DNA Pol III of *T. thermophilus* (Stano *et al.,* 2006) and *T. aquaticus* (Wing *et al.,* 2008), the PHP domain of DNA Pol X of *B. subtilis* (Baños *et al.,* 2008) and *T. thermophilus* (Nakane *et al.,* 2009) or the PHP domain of DNA Pol C of *S. pneumoniae* (Standish *et al.,* 2013). In contrast, neither the PHP domain of *G. kaustophilus* DNA Pol C presented exonuclease activity (Evans *et al.,* 2008). Aravind and Koonin (1998) had already suggested that the *E. coli* DNA Pol III PHP domain could be involved in pyrophosphatase activity. The tertiary structure of the PHP domain consists of an α/β barrel, containing seven β-strands surrounded by seven α-helices. One of the β strands runs in an anti-parallel sense, thus conferring a distorted conformation (Aravind & Koonin, 1998).

For the evaluation of the potential pyrophosphatase activity of the PHP domain of α subunit, the isolated PHP domain was previously found to feature this catalytic activity (Conte, 2012) (Fig. 4.3), according to the same coupled-enzyme assay (Suárez *et al.,* 2012) performed for the study of the polymerase and pyrophosphatase activities of the purified proteins in the present Thesis. In this type of assay, a molecule of pyrophosphate originated after the incorporation of a dNTP to the DNA strand during polymerization, is transformed into two molecules of phosphate because of the pyrophosphatase reaction. This phosphate is finally converted into uric acid after a series of coupled-enzyme reactions (Fig. 2.2). The changes in uric acid absorbance at 293 nm are used to determine the concentration of phosphate released as a result of the polymerase reaction and subsequent pyrophosphatase activity, assuming that the molar extinction coefficient (ε) of uric acid is equal to 12.5 x 10^3 M^{-1} cm^{-1} (Sheibe *et al.,* 1974). Following this method, a consistent pyrophosphatase activity was detected in the presence of the isolated PHP domain when 1 mM pyrophosphate was used as substrate (Conte, 2012) (Fig. 4.3), while in the absence of PHP or substrate the concentration of uric acid as a function of time did slightly increase, probably because of the presence of contaminating phosphate. In addition, first-order rate constant was equal to 0.2 s^{-1}, representing the hydrolysis of pyrophosphate by the isolated PHP domain.
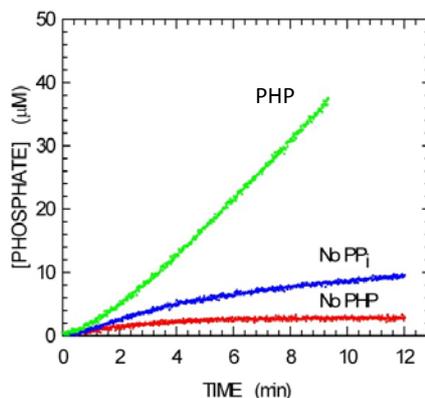
Figure 4.3: Pyrophosphatase activity assay of the PHP domain of α subunit (green dots). The pyrophosphatase activity was also tested in the absence of PHP domain (red dots) and in the absence of pyrophosphate (blue dots) (Conte, 2012).

Then, the purified α subunit was found to perform the pyrophosphatase activity under the same coupled-enzyme analysis. Besides, its pyrophosphatase activity was inhibited by sodium fluoride (Fig. 3.14), a specific inhibitor of pyrophosphatases (Baykov *et al.,* 2000) displacing the water molecule responsible for the nucleophilic attack and therefore impairing catalysis. The inorganic pyrophosphatase was used as control, showing zero-order kinetics in the absence of fluoride but with a clear deviation from this zero-order kinetics following the addition of sodium fluoride. k_{cat} was 103 and 0.33 s^{-1} for the inorganic PPase and the wild type α subunit, respectively.

Considering the concentrations of both assayed enzymes, it was noticed that the activity of inorganic PPase was about 180 times higher than that of α subunit. However, Mn^{+2} was omitted in these assays so as to avoid competition between fluoride and Mn^{2+}, as it was reported for the pyrophosphatase from *Streptococcus gordonii* (Parfenyev *et al.,* 2001). As it is below-discussed, Mn^{+2} promotes the development of the pyrophosphatase reaction in the PHP domain of α subunit. Due to the fact that Mn^{+2} is the metal activator of Family II PPases (Young *et al.,* 1998), whereas Mg^{+2} activates the pyrophosphatase reaction of type I PPases (Heikinheimo *et al.,* 2001), the pyrophosphatase activity of the PHP domain of α subunit of *E. coli* DNA Pol III is analogous to a type II pyrophosphatase activity.

Furthermore, a kinetic model of pyrophosphatase activity inhibition by NaF was devised. Enzyme inhibition by sodium fluoride is supposed to occur according to this scheme:

$$E + F \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E_0$$
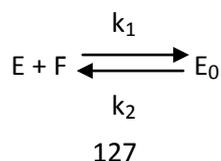
E represents the active enzyme, F represents fluoride ion and E_0 represents the inactive enzyme. NaF inhibits prokaryotic PPases in a rapid and reversible way (Baykov *et al.,* 2000). The kinetic model explains the pyrophosphatase activity development in the presence of inhibitor. The rate of inactive enzyme formation is given by the following equation:

$$\frac{d[E_0]}{dt} = k_1\,[F]\,[E] - k_2\,[E_0] \quad \text{which becomes} \quad [E_0] = \frac{[E_t]k_1\,[F](1-e^{-k_3t})}{k_3} \quad \text{after integrating.}$$

In this case, E_t represents total inorganic PPase enzyme whereas k_3 is equal to $k_1[F] + k_2$.

Considering NaF as a reversible inhibitor of the pyrophosphatase reaction, the following state is reached at equilibrium:

$$[E_0] = \frac{[E_t]k_1\,[F]}{k_3}$$

Assuming that the rate of reaction is given by:

$$\frac{[E_t]k_1\,[F]}{k_3} = k_{cat}\,([E_t] - [E_0])$$

the equation that represents the trend of the inhibition curve of PPase is:

$$[P] = \frac{k_{cat}[E_t]\,k_2}{k_3}\,t + \frac{k_{cat}[E_t]\,k_1\,[F]}{k_3}\,\frac{1}{k_3}\,(1-e^{-k_3t})$$

Tables 20A and 20B summarize k_1 and k_2 values for the different inhibition curves of the PPase (0.45 nM) and the wild type α subunit (30 nM), respectively, from graphics 3.14.A and 3.14.B (chapter 3), respectively, using increasing concentrations of NaF.

A	[NaF] (µM)	k_1 ($\mu M^{-1} s^{-1}$)	k_2 (s^{-1})
	20	10.43×10^{-6}	0.61×10^{-3}
	40	9.23×10^{-6}	1.14×10^{-3}
	100	14.85×10^{-6}	1.95×10^{-3}
	200	10.25×10^{-6}	1.67×10^{-3}
	400	7.10×10^{-6}	1.58×10^{-3}
	800	5.09×10^{-6}	7.3×10^{-4}
	1200	3.73×10^{-6}	6.4×10^{-4}

B	[NaF] (µM)	k_1 ($\mu M^{-1} s^{-1}$)	k_2 (s^{-1})
	20	5.45×10^{-4}	0.18×10^{-2}
	50	3.16×10^{-4}	0.19×10^{-2}
	200	7.26×10^{-4}	2.19×10^{-2}
	800	1.51×10^{-5}	0.086×10^{-2}
	1200	1.14×10^{-5}	0.0468×10^{-2}

Table 20A: k_1 and k_2 values for each concentration of NaF in the presence of 0.45 nM PPase; **Table 20B:** k_1 and k_2 values for each concentration of NaF using 30 nM α subunit.

The enzyme-Fluoride dissociation constant (K_D) was determined for enzyme-inhibitor complexes (PPase-Fluoride and α-Fluoride) according to the ratio between k_2 and k_1 and taking into account the reversible inhibition reaction. Fig. 4.4 represents the K_D values estimated as a function of inhibitor concentration. The K_D for the fluoride-PPase complex was equal to 159 ± 36 µM whereas the K_D for the fluoride-α subunit complex was 28 ± 23 µM. Therefore, α subunit was much more inhibited by the action of NaF than the inorganic PPase.
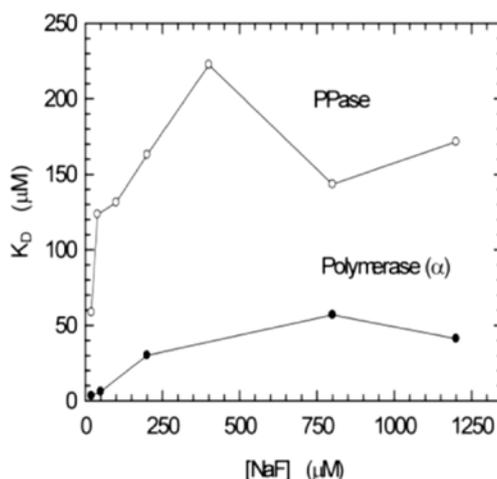
Figure 4.4: K_D values for inorganic PPase and α subunit as a function of sodium fluoride concentration.

Taking into account the equation at the equilibrium, above mentioned, the residual pyrophosphatase activity of both PPase and α subunit was also calculated. Results are summarized in Table 21A and Table 21B. This value represents the concentration of inactive enzyme present in the reaction mixture after NaF inhibition, assuming that the equilibrium for the enzyme-fluoride complex was reached after 800 seconds in the case of the inorganic PPase, and after 500 seconds for α subunit, obtaining a steady-state of the fluoride-enzyme complex. The amount of enzyme that is still active derives from the difference between

the total known enzyme concentration and the calculated concentration of inactive enzyme. The determined residual activity was in good agreement with the value calculated with the non-linear kinetic model, except for a minimum error derived from the experimental procedure.
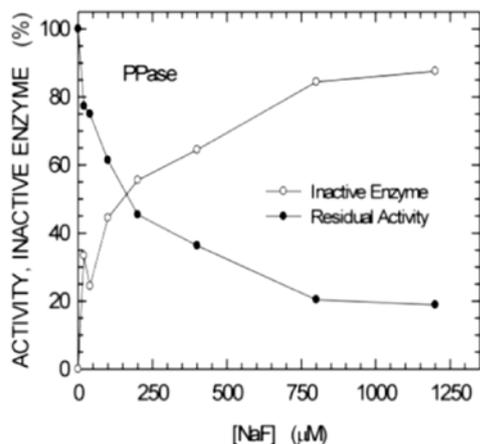
[NaF] (μM)	[E$_0$] (μM)	Residual activity (μM/s)
0	0	0.044
20	$0.15 \cdot 10^{-3}$	0.034
40	$0.11 \cdot 10^{-4}$	0.033
100	$0.20 \cdot 10^{-3}$	0.027
200	$0.25 \cdot 10^{-3}$	0.020
400	$0.29 \cdot 10^{-3}$	0.016
800	$0.38 \cdot 10^{-3}$	0.009
1200	$0.39 \cdot 10^{-3}$	0.0083

Fig. 4.5.A: Residual activity of the inorganic PPase as a function of NaF concentration; **Table 21A:** values of inactive PPase and residual activity for each concentration of NaF.
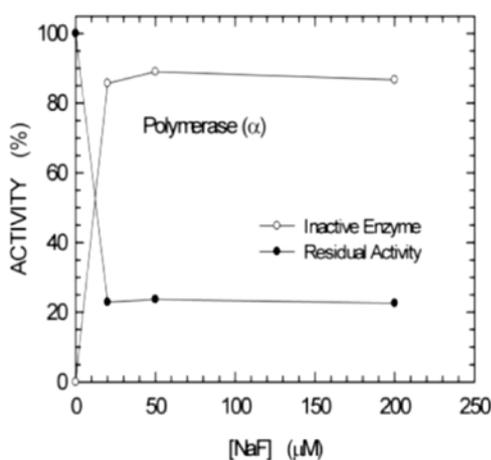
[NaF] (μM)	[E$_0$] (μM)	Residual activity (μM/s)
0	0	0.01000
5	$2.75 \cdot 10^{-2}$	0.00249
20	$2.57 \cdot 10^{-2}$	0.00226
50	$2.67 \cdot 10^{-2}$	0.00229
200	$2.60 \cdot 10^{-2}$	0.00237
800	$2.79 \cdot 10^{-2}$	0.00111
1200	$2.89 \cdot 10^{-2}$	0.00104

Fig. 4.5.B: Residual activity of the α polymerase as a function of NaF concentration; **Table 21B:** values of inactive α polymerase and residual activity for each concentration of NaF.

The characterization of the PHP domain as a type II pyrophosphatase made select some type II inorganic PPases (from *Streptococcus mutans* and *Bacillus subtilis*) for a comparative study, with the purpose of looking for the catalytic amino acids of α subunit involved in pyrophosphate hydrolysis. The tertiary structure of the reported wild type α subunit (Lamers *et al.,* 2006) was used for the study of the candidate active site, paying particular attention to the PHP region in which the inorganic phosphate was found. This

phosphate could represent the reaction product of the pyrophosphatase reaction. In both inorganic PPases from *S. mutans* and *B. subtilis*, a cluster of four aspartates and a couple of histidines is similarly oriented in the active site. By sequence homology, H12, D19, D43, D69, H83, and D201 were considered as members of the *E. coli* PHP domain active site and, consequently, three site-specific mutants (H12A, D19A, and D201A) were constructed.

The aspartate D201 side chain points towards the phosphate and the putative metal coordination site, interacting with the negative charges of the phosphate at the interface of the solvent. Presumably, D201 may play a similar role to that of performed by the D117 of yeast inorganic pyrophosphatase (Heikinheimo *et al.,* 2001), responsible for water molecule activation. On the other hand, the side chain of H12 and D19 residues points towards the putative P1 phosphate direction, meaning that both could be involved in coordinating metal ions. The wild type and mutant *dna*E genes were cloned into the pBAD vector, and protein overexpression was reached by addition of arabinose to the culture medium, obtaining high expression levels of each mutant. However, the concentration of the αD201A was rather low in the soluble fraction of protein extracts and only SDS (0.1%) was able to increase its solubility, in contrast with other agents (Fig. 3.2). However, since SDS is negatively charged and could interfere during the anion exchange purification step, αD201A protein was preferred to be assembled to τ, ε and θ proteins. Therefore, while the H12A and D19A variants were purified individually, the αD201A variant was co-expressed in *E. coli* with τ, ε, and θ subunits by a dual-plasmid system based on the pBAD and pGOOD plasmids. The ατD201Aεθ complex was obtained, avoiding problems related to poor solubility or partial degradation as for the single αD201A protein.

Pyrophosphatase activity assays were performed by using 1 mM pyrophosphate as substrate, in the presence of 10 and 0.25 mM of $MgCl_2$ and $MnCl_2$, respectively. 17 nM α subunit showed an activity rate equal to 21 nM/s (Fig. 3.13.A). Under the same conditions the activities observed in the presence of 17 nM H12A or 13 nM D19A were 2.1 and 3.8 nM/s, respectively (Fig. 3.13B). Subsequently, the corresponding first-order rate constants were equal to 1.23 s^{-1} for α wild type, 0.12 s^{-1} in the case of H12A and 0.2 s^{-1} for D19A. These observations unequivocally determine the main role of H12 and D19 in pyrophosphatase activity catalyzed by the PHP domain, since their substitution by alanine impaired the hydrolysis of pyrophosphate 10 and 5.5 times, respectively, when compared to wild type α subunit.

On the other hand, reaction velocities were 11 and 63 nM/s for the wild type ατεθ complex and the mutant αD201Aτεθ complex, respectively; first-order rate constant was equivalent to 2.4 s^{-1} in the case of the ατεθ complex, and 14 s^{-1} for the αD201Aτεθ variant. Surprisingly, the mutation at position 201 of α subunit was

found to increase the catalytic efficiency. As it was previously mentioned, the D201 residue is close to the phosphate ion located at the PHP domain of α subunit. Therefore, this D201 residue would have the role of allowing the release of phosphate, the reaction product, and the substitution of this aspartate with alanine would considerably favour this release, avoiding product inhibition of the catalytic activity of the PHP domain. In the case of the inorganic PPase of *E. coli*, another aspartate residue (D67) was found to be responsible for promoting the release of a phosphate molecule by modifying the disposition of its carboxylate group (Samygina *et al.,* 2007).

In addition, the polymerase activity of the wild type α subunit, its mutant variants H12A and D19A and both αεθ and αD201Αεθ complexes was also analyzed, in the absence and in the presence of inorganic PPase and by the same coupled-enzyme assay. In the absence of exogenous PPase, the determination of uric acid would be exclusively related to the pyrophosphatase activity of the PHP domain, as a consequence of the polymerase reaction; in contrast, the presence of an excess of inorganic PPase would indicate the maximal rate of DNA extension. By confronting the activity in the presence and absence of inorganic PPase, the degree of coupling between DNA elongation, performed by α subunit, and the intrinsic pyrophosphatase activity of the PHP domain can be determined.

The wild type α subunit (17 nM) was tested in the presence of 1 μM DNA, observing reaction velocities equal to 17.3 and 24.8 nM/s in the absence and presence of inorganic PPase, respectively. Subsequently, the polymerase activity in the absence of PPase was 22% lower than that of using 1 mM pyrophosphate as substrate (21 nM/s; Fig. 3.13) and a 30% lower than that of including 1 μM DNA (24.8 nM/s; Fig. 3.15) in the presence of PPase. These observations suggest that the coupling between both DNA elongation and pyrophosphatase activities is partial (70%) in *E. coli* DNA Pol III α subunit. This coupling between the two catalytic activities was considerably lower for αH12A (17 nM) and αD19A (13 nM) proteins. In the absence of inorganic PPase, the reaction velocities were 1.8 and 2.9 nM/s for H12A and D19A, respectively, whereas in the presence of PPase these speeds were equal to 4.9 for αH12A, and 8 nM/s in the case of αD19A. For both mutants, the coupling between pyrophosphatase and polymerase activities was equal to 36%, a half respect to the wild type α subunit, and this inefficient coupling is consistent with the decrease in the maximal rate of DNA elongation. Concerning the αεθ complex (5 nM), the reaction velocity was 14.7 nM/s in the absence of PPase, and 19.9 nM/s in the presence of inorganic PPase. The corresponding velocities for the αD201Αεθ (3.2 nM) were 24.5 and 13.2 nM/s, with and without exogenous PPase, respectively, thus confirming that the mutant complex develops faster polymerase and pyrophosphatase activities than the wild type complex. Actually, the coupling between both activities was equal to 73.8% and 53.9% for the wild type and the mutant complex, respectively.

Overall, site-specific substitutions influenced the polymerase activity in two different ways: a sharp decrease in polymerase activity was observed for the αH12A and the αD19A mutants concluding that, although not directly involved in the polymerase reaction, these two amino acids significantly affect the development of DNA elongation; in contrast, the αD201A variant features a two-fold increase in the polymerase activity when compared to the ατεθ complex. However, while site-specific mutagenesis has provided interesting information for defining the catalytic role of those residues subjected to mutation, the conformation of the entire active site of the PHP domain is still incomplete. Nevertheless, there are clear similarities between the catalytic sites of the pyrophosphatase II of *B. subtilis* or *S. mutans* and the PHP domain of the DnaE polymerase of *E. coli*, as it was previously mentioned (Fig. 1.19). It is noteworthy the presence of two sets of amino acids involved in distinct catalytic functions. The histidine 12 and the aspartate 19 are part of the group of amino acids responsible for the coordination of metal ions. On the other hand, aspartate 201, present at the interface with the solvent, would interact with the negative charges of phosphate molecules and would control its release after the hydrolysis of the pyrophosphate.

Moreover, the dependence on manganese of the pyrophosphatase activity (Fig. 4.6) further emphasizes the similarity between the PHP domain and Family II pyrophosphatases, since only pyrophosphatases type II, which present two histidines involved in metal coordination, are activated by the presence of manganese (Young *et al.,* 1998), in contrast with the Mg^{+2} requirement for type I PPases (Heikinheimo *et al.,* 2001). Pyrophosphatase activity assays were performed with a 1 mM inorganic pyrophosphate and a low concentration of manganese (250 µM), being essential for activating the hydrolysis of pyrophosphate. This low ion concentration highlights an internal mechanism of transit from the polymerase to the pyrophosphatase enzyme.
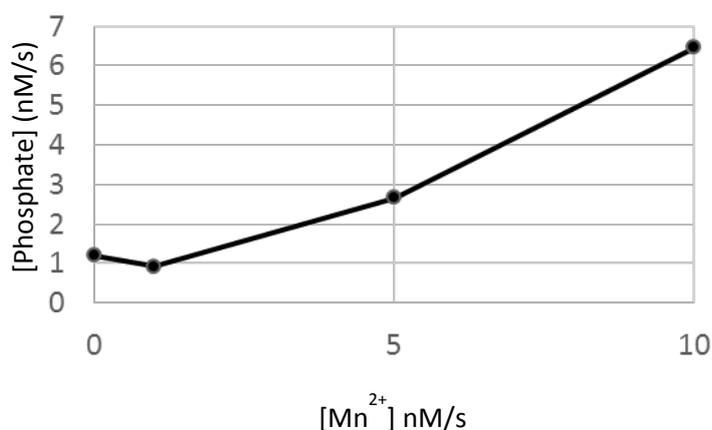
Figure 4.6: Polymerase activity in the absence of exogenous pyrophosphatase. Assays were performed with 17 nM α subunit, 10 mM MgCl2 and different concentrations of MnCl2 (0, 5 and 10 mM). As we are evaluating polymerase activity in the absence of PPase, we are indeed studying the role of the PHP domain of α subunit on hydrolyzing the pyrophosphate generated as a result of nucleotide incorporation. This reaction depends on the concentration of Mn^{+2}.

The amount of pyrophosphate produced by the polymerization reaction that the wild type PHP domain is incapable of hydrolyzing seems to be about 1/3 of the total, since the coupling between the polymerase and pyrophosphatase activities is not fully complete (70%). Owing to this coupling, there must be some type of structural element connecting both polymerase and pyrophosphatase active sites. A pyrophosphate channel connecting the palm (where pyrophosphate is generated) and the PHP domain may be hypothesized (Fig. 4.7) and, subsequently, the pyrophosphate not channeled to the PHP domain would be released in solution, being thus trapped and hydrolyzed by the inorganic PPase, if present. The H12A and D19A site-specific mutant variants decreased the level of coupling (35%) between the two catalytic reactions, as it was seen from the polymerase assays in the presence and in the absence of exogenous pyrophosphatase, due to the fact that these mutants were not able to coordinate metal ions properly. The accumulation of pyrophosphate in these mutant variants acts as a product inhibitor of the polymerase reaction.
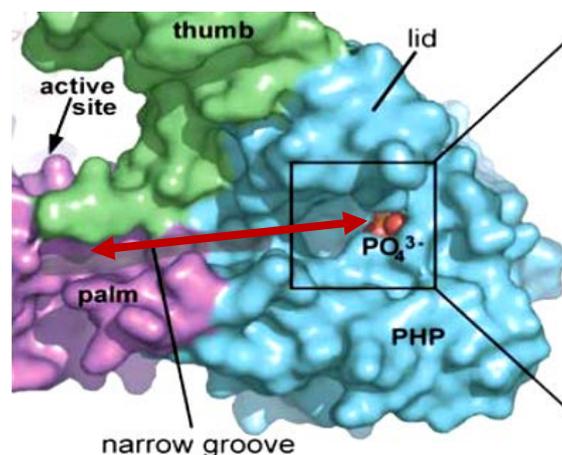
Figure 4.7: Surface representation of DNA Pol III α subunit, highlighting the PHP (in blue), the Thumb (in green) and the Palm (in purple) domains (Lamers *et al.,* 2006). A molecule of phosphate was found in the putative active site of the PHP domain. A harrow represents the groove that runs from the Palm domain into a cavity of the PHP domain in where the phosphate molecule was found. Towards this groove, the molecule of pyrophosphate could go from the polymerase active site to the PHP domain active site in order to be hydrolyzed by the pyrophosphatase activity of the PHP domain.

Regarding the trimeric complexes, the situation concerning the polymerase activity of the αεθ complex was similar to that of the single α subunit. However, the pyrophosphatase activity was significantly different, since the PHP domain of the αεθ complex did not seem to have direct access to the pyrophosphate in solution, probably because of the type of complex conformation. Actually, the trimeric complex should have shown activity three times higher than that of α subunit, but results showed that this

activity was only twice respect to the single polymerase. Therefore, the ατεθ complex lost a third of the expected activity. On the other hand, the αD201Aτεθ mutant showed a high ability for hydrolyzing pyrophosphate in solution. For this reason, despite the fact that the association between α and τεθ subunits disadvantaged the spread of pyrophosphate from solution to the PHP domain, the D201A substitution enabled the complex to stimulate pyrophosphatase activity and the subsequent release of phosphate product.

The coupling between polymerase and pyrophosphatase activity could also be considered as a regulatory system responsible for controlling the rate of DNA elongation. Consequently, the modification of this coupling may well directly confer a phenotype to *E. coli* cells and populations. To test the effect of α subunit mutant variants on *E. coli* phenotype, αD201A and αH12A were overexpressed and cell morphology, viability and the mutation frequency of cells hosting these variants were compared with those of cells overexpressing the wild type α subunit. The overexpression of both mutant variants, but specially the αD201A, resulted in the presence of aberrant phenotypes, characterized by longer cells bearing multiple nucleoids owing to the difficulty for those cells to complete binary fission (Fig. 3.24). The amount of aberrant cells was much higher for cells bearing the D201A and for the H12A protein than for the wild type protein individuals. Moreover, severe growth defects were observed as a result of the induction of αD201A (Fig. 3.22.A) whereas the αH12A practically achieved the same population density than after overexpressing the wild type α subunit (Fig. 3.22.B), although cell mortality after 8 hours growing at 37° C was even higher for H12A than in the case of the D201A variant (Fig. 3.23). Finally, the frequency of appearance of resistant individuals against rifampicin in the different populations was tested in order to determine whether the defective phenotypes conferred by α subunit variants also corresponded to an increase in mutation rates. The mutation frequency of *E. coli* populations induced or not to the overexpression of α subunit was $6.7\pm5.9\times10^{-8}$ and $1.4\pm0.4\times10^{-8}$, respectively. On the contrary, when both not induced and induced cultures of *E. coli* hosting the αD201A or the αH12A variant were evaluated, resistant clones presented an apparent 10^{-2} frequency after culturing both mutants for a few generations. However, this mutation frequency is so high that quantifying its exact magnitude was impossible. This mutational level is indeed consistent with a lethal mutation rate (Painter, 1973; Ninio, 1991) and, therefore, is not compatible with cell viability.

The correlation between the pyrophosphatase and the polymerase activities makes the PHP domain a suitable new target for novel antibiotics. Nowadays, drug resistances is a serious challenge to global public health that requires a deep research with the objective of determining potential new cell elements to be specifically inhibited. Standish and co-workers (2013) found that the fascioquinol E selectively inhibited the

protein tyrosine phosphatase CpsB and the DNA polymerase PolC of *S. pneumoniae*, being both proteins member of the PHP family. After DNA PolC inhibition, the growth of Gram positive population was remarkably inhibited, even though fascioquinol E did not affect Gram negative bacteria. In this Thesis, cell viability was clearly altered after both decreasing (D19A and H12A variants) and increasing (D201A variant) the pyrophosphatase activity of the PHP domain of α subunit, due to the alteration of the coupling between polymerase and pyrophosphatase activity. Therefore, the PHP domain of the DNA Pol III of *E. coli* is a potential new target for designing specific inhibitory molecules that could arrest Gram negative bacteria growth.

4.3 HoLaMa: an artificial mini DNA polymerase

HoLaMa is a catalytically active mini-DNA polymerase designed by protein engineering (Martina, 2014), in order to understand more efficiently the mechanism of reaction of DNA polymerases. The artificial DNA polymerase features the following three properties:

- Presence of 5' → 3' polymerase activity exclusively, without exonuclease activity, unlike the Klenow Fragment.
- Small molecular mass (46 kDa), suitable for nuclear magnetic resonance (NMR) analysis.
- Ability for fluorescent molecules binding at pre-determined sites.

HoLaMa enzyme was designed by molecular modelling from the tertiary structure of the Klenow polymerase, filed on *RCSB Protein Data Bank* (Brautigam *et al.*, 1999). The structure of reference is "*2KZZ*", and three crystal structures were also considered (Beese *et al.*, 1993; Teplova *et al.*, 1999; Brautigam *et al.*, 1999). The molecular engineering was pursued using different softwares:

- Visual Molecular Dynamics (VMD), (Humphrey, 1996).
- Swiss PDB Viewer (Deep View), (Guex, 1997).
- Foldit-Standalone (University of Washington, 2008).

Two variants of HoLaMa enzyme were designed. The difference between them was the presence (HoLaMa-his) or the absence (HoLaMa) of a hexa-histidine tag at the N-terminus. Nevertheless, only the overexpression, purification and characterization of HoLaMa devoid of His-tag was performed in the present thesis, whereas Martina (2014) was responsible for characterizing the one bearing the histidine tag (HoLaMa-his). The presence of a histidine tag enables the selective purification with metal ion affinity

chromatography. Due to removing the 3' → 5' exonuclease domain (M324-M519), a hydrophobic G521-N543 α-helix was exposed to the solvent, while in the Klenow enzyme is placed as an internal interface between 5' → 3' polymerase and 3' → 5' exonuclease domains. Thus, eleven site-specific mutations were introduced, taking into account the variation of the scoring function associated to any single change by considering results provided by Deep View and Foldit-Standalone softwares (Fig. 4.8.A). K520M, the first amino acid of the protein sequence was turned into methionine to maintain standard rules of mRNA translation; P522E, L523K, F526E, M531R, V534E and P535R mutations were performed to optimize the exposed hydrophobic G521-N543 α-helix and increase protein solubility; L561K and L646D, both exposed to the solvent, were included to achieve the lowest energy state; L744C and C907S introduced a single fluorophore binding-site. Therefore, HoLaMa was the result of Klenow fragment engineering by 11 site-specific mutations and deletion of 195 amino acids (409 amino acids respect to the 928 amino acids of the wild type *polA* gene) (Fig. 4.8.B).
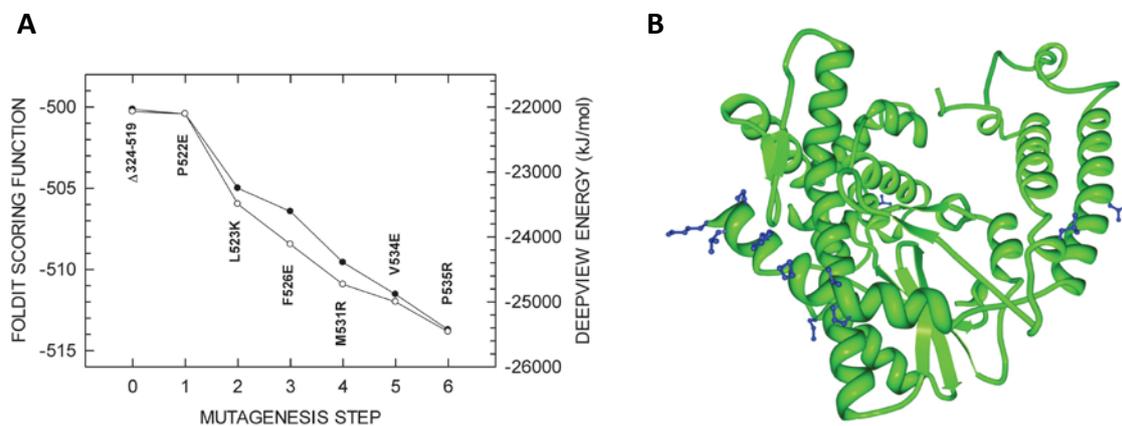
Figure 4.8: (A) Variation of DeepView (empty circles) and Foldit (filled circles) scoring functions as a result of site-specific mutations in the G521-N543 α helix of the Klenow sub-fragment comprising residues 520-928 of *E. coli* DNA Pol I; **(B)** Partial tertiary structure of Klenow fragment (PDB 2KZZ). The region containing residues 520-928 is reported in green; 11 site-specific substitutions are shown in blue.

HoLaMa sequence was synthesized (Entelechon GmbH, Regensburg, Germany) and cloned into the pBAD plasmid (Invitrogen) with PstI and NcoI. Protein overexpression was triggered with arabinose for 15 hours at 30° C. Problems related to low protein expression or solubility were not found in contrast with other DNA Pol I engineering (Freemont *et al.,* 1986), meaning that mutations performed for stabilizing the exposed α-helix were rather successful. The purification protocol for HoLaMa included an anion exchange Q-Sepharose, an affinity Cibacron blue and a final Hitrap heparin column. A His-trap affinity column was not used, due to the absence of a histidine tag. Only the first anion exchange step was significantly different when compared to the purification procedure of HoLaMa-his. The elution of HoLaMa devoid of the

histidine tag occurred during the washing step with buffer containing 50 mM NaCl, before the NaCl gradient started; on the contrary, the elution of the HoLaMa-his variant was found at a NaCl concentration of about 260-280 mM. Therefore, the binding to the anion exchange matrix was weaker in the case of HoLaMa devoid of the histidine tag, presumably because of an altered folding in the absence of histidine tag or the influence of the tag on exposing more negatively charged residues to the solvent. No significant differences were found during the Cibacron blue or the Hitrap heparin chromatographic steps. Both the expression level and the final protein yield were slightly lower in the case of HoLaMa (0.2 mg from 6 litres of culture) when compared to HoLaMa provided with the histidine tag (0.9 mg from 6 litres of culture).

The reason of purifying two different variants of this artificial mini DNA polymerase was to determine whether the presence of the histidine tag was responsible for impairing the activity of the protein. Actually, k_{cat} of HoLaMa-his was considerably lower than that of Klenow enzyme (Martina, 2014). When 1.5 nM Klenow enzyme and 51 nM HoLaMa or HoLaMa-his were used, a K_m equal to 0.77 µM, 1.85 µM and 1.1 µM was found in the case of Klenow enzyme, HoLaMa and HoLaMa-his, respectively (Fig. 4.9; Table 22); therefore, the K_m was 2.4 (HoLaMa) and 1.43 (HoLaMa-his) times higher than the corresponding constant of Klenow enzyme. The K_m of the Klenow Fragment C terminus (Derbyshire et al., 1993) was 1.4 µM, in good agreement with that of HoLaMa. Under the same conditions, Klenow enzyme presented a k_{cat} equivalent to 8 s^{-1} while those of HoLaMa and HoLaMa-his were 0.31 and 0.06 s^{-1}, respectively. HoLaMa and HoLaMa-his showed 0.5 and 1.6 % catalytic efficiency, respectively, when compared with Klenow enzyme, at least, under steady-state conditions. Table 22 summarizes kinetic data.
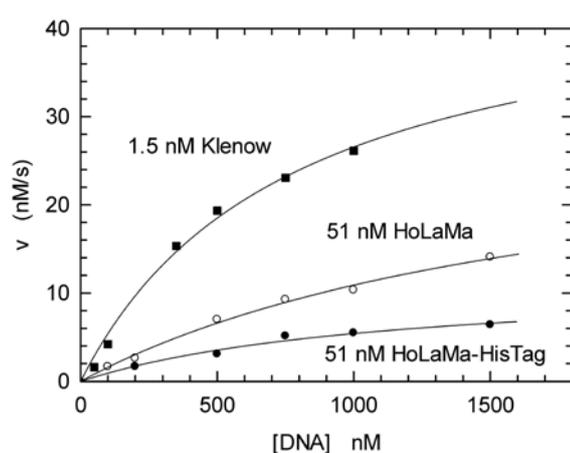
	Klenow	HoLaMa	HoLaMa-his
K_m	0.77 µM	1.85 µM	1.1 µM
k_{cat}	8 s^{-1}	0.31 s^{-1}	0.06 s^{-1}

Table 22: Kinetic data obtained for Klenow enzyme (1.5 nM), HoLaMa (51 nM) and HoLaMa-his (51 nM).

Figure 4.9: Kinetics of processive DNA polymerase activity as a function of DNA concentration, using 1.5 nM Klenow enzyme (filled squares), 51 nM HoLaMa (empty circles) and 51 nM HoLaMa-His (filled circles). Continuous line show best fits to the Michaelis-Menten equation. 1 µM 40mer DNA and 100 µM dTTP were present in polymerase reactions.

It is thus clear that the histidine tag interacts with part of the protein impairing the polymerase activity. Remarkably, k_{cat} of HoLaMa was considerably higher than the catalytic constant previously defined for a Klenow sub-fragment (0.0012 s-1, Derbyshire *et al.,* 1993). The increase of K_m was significant, compared with Klenow enzyme, and it is likely related to the fact that HoLaMa and HoLaMa-his lack the 3' → 5' exonuclease domain. This proofreading domain may well be responsible for DNA disposition during DNA binding despite not participating actively in polymerization. Actually, Beese and co-workers (1993) found that DNA comes into the Klenow enzyme towards a cleft located between the polymerase and the 3' → 5' exonuclease domains. Mutant Taq Pol I in which the 5' → 3' was absent showed a processivity and fidelity of replication significantly lower than those of the wild type Taq Pol I (Merkens *et al.,* 1995). Hence, although not directly involved in polymerase activity, the exonuclease domain of DNA polymerases is quite important for the correct DNA elongation.

The exonuclease domain may also be important to keep the polymerase active site in a proper conformation. From Klenow Fragment tertiary structures (Beese *et al.,* 1993; Teplova *et al.,* 1999; Brautigam *et al.,* 1999) it is noted the presence of an α-helix between the polymerase and exonuclease domains. That helix is provided with two prolines, probably necessary to maintain the rigid structure in a proper conformational arrangement. Both HoLaMa and HoLaMa-his lack these two proline residues so that, in addition to the deficient exonuclease domain, the α-helix may be more relaxed, in contrast to the Klenow polymerase, impairing the conformation of the polymerase active site and, subsequently, the activity of the artificial enzyme. In addition, this more relaxed conformation could trigger a slowed down running of the primer terminus along the space located between the polymerase and exonuclease domains.

Different strategies have been developed for increasing the processivity and fidelity of replication in DNA polymerases, especially in those ones lacking the exonuclease domain or other elements whose absence is directly related to an impairment of the enzyme activity. These strategies are focused on using proteins or sequences which are known to improve processivity (processivity factors). Including a fused sequence of a double-stranded DNA binding protein (Sso7d) from *Sulfolobus solfataricus* was found to be an optimum candidate to substitute Taq Pol I exonuclease domain and increase enzyme processivity and accuracy without altering protein activity or stability (Wang *et al.,* 2004). Also the fusion of helix-hairpin-helix motifs from DNA topoisomerase V increased processivity in different polymerases, such as the Taq Pol I lacking exonuclease domain (Pavlov *et al.,* 2002) or the φ29 DNA polymerase (de Vega *et al.,* 2010). In addition, the binding of thioredoxin increased remarkably the processivity of *E. coli* DNA Pol I upon fusing the domain from T7 DNA polymerase, responsible for binding thioredoxin, into the sequence of DNA Pol I (Bedford *et*

al., 1997). Mutations performed in highly conserved Motif I in Family A polymerases tend to be a suitable strategy to increase processivity (Patel & Loeb, 2000; Reha-Krantz *et al.,* 2014). However, the increase of processivity in some cases could result in a severe decrease of accuracy during DNA replication, especially important when considering the absence of the exonuclease domain.

Unexpectedly, HoLaMa devoid of the histidine tag presented an initial velocity more than three times higher (9.4 nM/s) than that of Klenow enzyme (3 nM/s) when the polymerase assay was performed under non-processive conditions (Fig. 4.10). On the contrary, initial enzyme activity of HoLaMa-his appears to be linear and three times lower (1.1 nM/s) than that of Klenow enzyme (3 nM/s) under the same conditions (Fig. 4.10). This assay was carried out using a dsDNA generated from the annealing of a 16mer template to a 15mer primer, thus allowing the extension of a single nucleotide. The catalytic efficiency in the absence of processivity is an interesting parameter to compare the ability of Klenow enzyme and these artificial polymerases to bind the DNA, add one nucleotide, dissociate from DNA and repeat the cycle. Klenow enzyme is capable of adding about 12 nucleotides without dissociating from the DNA (Kong *et al.,* 1993). HoLaMa may dissociate DNA very quickly probably because of a non-proper conformation.
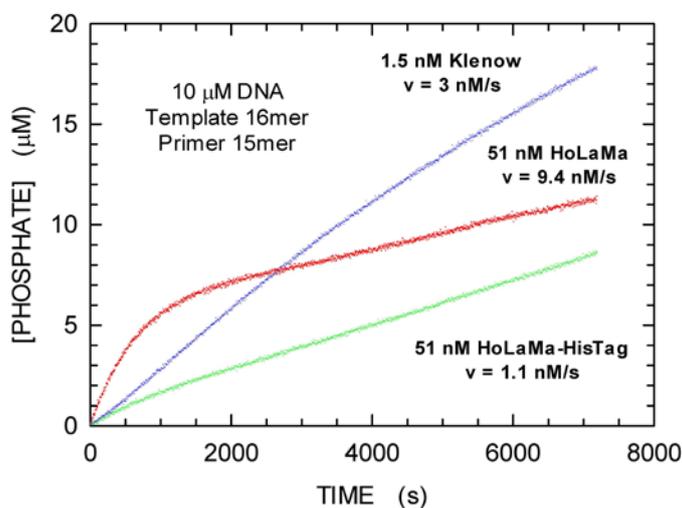
Figure 4.10: Enzyme kinetics in the absence of processivity for 1.5 nM Klenow enzyme (blue-dotted line), 51 nM HoLaMa (red-dotted line) and 51 nM HoLaMa-His (green-dotted line) using 10 µM 16mer dsDNA and 100 µM dTTP.

The k_{cat}/K_m for the non-processive reactions was 0.027 and 0.061 $\mu M^{-1} s^{-1}$ for HoLaMa and Klenow enzyme, respectively. It was again observed that the histidine tag did negatively affect polymerase activity. Further studies are required to determine the mechanism by which exonuclease domains affect properties of DNA polymerases.

4.4 Study of DNA polymerases by single-molecule FRET

Single-molecule Förster Resonance Energy Transfer (smFRET) enables the determination of distances between two fluorophores, within a range from 2 to 10 nm. FRET is based on a non-radiative energy transfer from a donor fluorophore to an acceptor chromophore (Förster, 1946). In addition to the distance between dyes, the overlap of the donor emission and acceptor absorption spectra is required, and the relative orientation of the donor and acceptor transition dipole moments should enable transfer of energy (Lakowicz, 2006).

SmFRET has become an important tool for the study of DNA replication (Berezhna *et al.,* 2012) and transcription (Wang *et al.,* 2005), RNA folding (Keller *et al.,* 2014), DNA-protein interactions (Farooq *et al.,* 2014) or protein-protein interactions (Llères *et al.,* 2007). Importantly, smFRET allows us to resolve dynamic and static sample heterogeneity, which is often impossible in conventional biochemical analysis. Therefore, smFRET is ideally suited for studying molecular interactions and dynamics *in vitro* (Weiss, 1999; Hohlbein *et al.,* 2010). In addition, *in vivo* studies in combination with smFRET, were presented by Crawford and co-workers (2013), who observed single externally labelled biomolecules in living cells by combining smFRET imaging and electroporation.

A few limitations of FRET have to be taken into consideration before designing experiments:

- SmFRET requires the attachment of, at least, two extrinsic dyes to the molecules of interest. Protein labelling is not very easy and not always labelling rate is high.
- Fluorescence properties and energy transfer depends on the environment and orientation of the dyes.
- The number of photons that can be expected from a single organic fluorophore is limited. Therefore the use of oxygen scavenger systems becomes essential to guarantee a suitable time for imaging, avoiding photobleaching. Besides, time resolution in smFRET imaging does also depend on the frame rate of the emCCD camera.
- For observing isolated molecules, the background fluorescence has to be kept low; the highest concentration of labelled molecules freely diffusing in the sample volume must be smaller than 1 nM for confocal microscopy and smaller than 50 nM for TIRF microscopy.

- The necessity of controlling the concentration of fluorophores to avoid background and identify a molecule of interest instead of other similar molecules limits interactions between immobilized molecules. We need to considerer that normally high concentrations of molecules are required for studying some interactions due to the low dissociation constant (Levene *et al.,* 2003).

The use of modified or labelled DNA that can be easily immobilised on a modified glass surface allows the analysis of polymerase-DNA interactions and DNA polymerization by smFRET (Farooq *et al.,* 2014). After immobilization, DNA acts as a binding sensor for polymerases, present in solution, and DNA interactions can be studied at physiologically relevant conditions. Different techniques have been developed in order to allow the visualization of immobilized single-molecules, allowing the study of molecular dynamics upon positioning the laser over the molecule (Moerner & Fromm *et al.,* 2003). In 2009 the combination of smFRET and TIRF microscopy was used to visualise DNA synthesis (Christian *et al.,* 2009). A DNA template was labelled with a donor dye while the Palm domain of Klenow fragment was labelled with an acceptor probe. With this method, DNA-polymerase dynamics were detected (Fig. 4.11).
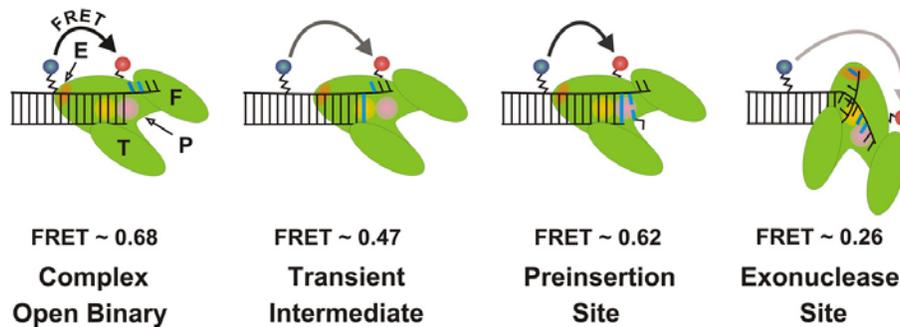
Figure 4.11: Models for DNA-polymerase dynamics by using FRET and by determining changes in energy transfer upon conformational change. The donor dye (blue sphere) is located on the DNA template, whereas the acceptor probe (red sphere) is on the exonuclease domain of *E. coli* DNA Pol I (Christian *et al.,* 2009).

Moreover, the opening and closing of the Fingers domain, essential during dNTPs incorporation, was observed (Minnick *et al.,* 1996). A change in the distance between the dyes, and the corresponding change in energy transfer, was related to DNA reorientation or DNA polymerization, dNTPs incorporation and even proofreading activity phenomena.

In addition, conformational changes can be predicted after observing different levels of FRET in a DNA polymerase labelled with a donor attached to the fingers domain and an acceptor attached to the Thumb domain: a lower FRET efficiency indicated an open conformation of Fingers domain with binary DNA-Pol I complexes; a higher FRET efficiency was preferentially found in the closed conformation associated to DNA-Pol I–dNTP ternary complexes (Hohlbein *et al.,* 2013). Nevertheless, Fingers domain did not close completely in ternary complexes with mismatches or when ribonucleotides were present (Santoso *et al.,* 2010; Hohlbein *et al.,* 2013) indicating the existence of a novel, partially close (*ajar*) state. Actually, a crystal structure of *Bst* Pol I (a structural homologue of *E. coli* DNA Pol I) bound to DNA was published (Wu & Beese, 2011), revealing the presence of an *ajar* conformation when Fingers domain binds to an incorrect nucleotide. Polymerase-DNA-dNTP complex exists in equilibrium among open, *ajar* and closed conformational states (Berezhna *et al.,* 2012). Therefore, smFRET enables us to detect FRET changes that accompany DNA/polymerase conformational changes and the binding of complementary or non-complementary dNTPs.

Unlike DNA Pol I and Klenow Fragment, which has been extensively studied by smFRET, DNA Pol III of *E. coli* has barely been studied. Shapiro *et al.* (2005) described a novel FRET assay for DNA polymerase activity of the DNA Pol III α subunit of *Haemophilus influenza.* Polymerase activity was related to an increase in FRET. The elongation of the primer supposed the production of a non-denaturable product. The increase of FRET between dyes was thus because of the resistance of the elongated product after polymerase activity of the assayed α subunit. This study represented one of the first approaches to our research, being one of the first attempts to analyze the DNA Pol III α subunit by FRET experiments, albeit not at the single-molecule level.

Furthermore, single-molecule techniques were used for analyzing some aspects of *E. coli* DNA Pol III. That was the case of the analysis of the advantages of an *in vitro* sub-assembled trimeric replicase, in comparison to a classic dimeric replisome, on lagging-strand synthesis and enzyme processivity (Georgescu *et al.* 2013). The study showed that single-stranded DNA gaps appeared on the lagging strand by the activity of the dimeric replisome, not completing the extension of Okazaki fragments. The dimeric replisome featured an impaired processivity owing to the fact that only in the case of a trimeric replisome the dissociated polymerase can be replaced by the third polymerase. Moreover, the study of the different phases of lagging strand synthesis by *E. coli*

replisome and how lagging-strand replication influences replisome processivity and speed of replication have also been studied at the single-molecule level (Georgescu *et al.,* 2014). However, some important aspects, such as the characterization of the single α polymerase of *E. coli* or the stimulating effect of ε subunit on polymerase activity, have not been determined at the single-molecule level. In addition, experiments combining single-molecule and FRET techniques are scarcely found in the literature of the DNA Pol III of *E. coli*.

As described in chapter 2 (2.2.24), DNA constructs were labeled with both donor and acceptor dyes, separated by a different number of bases depending on the sequence. DNA molecules were sparsely immobilized on a glass surface to guarantee single-molecule analysis. Hence, in our assay, an initial *E** was expected for each immobilized molecule. It was unknown whether and how *E** would change after the polymerase binds to the free 3' end of the DNA substrate. On the other hand, a decrease in FRET efficiency was expected for the polymerization process because of the increase in the distance between dyes upon DNA polymerization (Fig. 4.12).
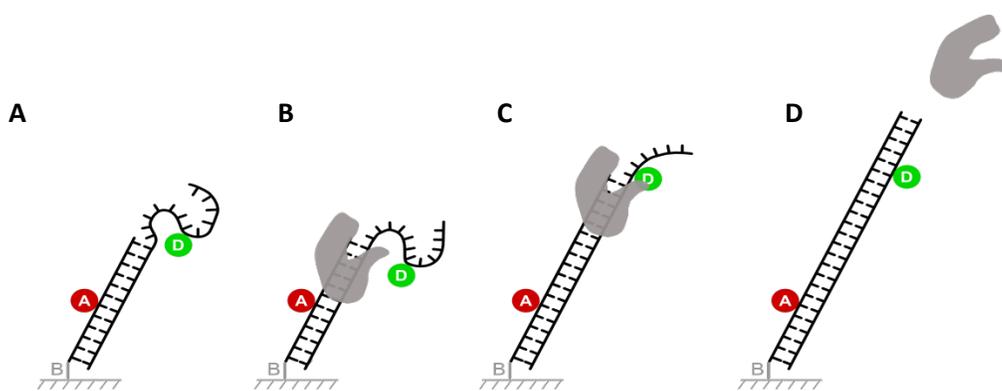
Figure 4.12: Graphical representation of DNA-construct conformational changes from initial FRET state (A), after protein binding (B), during DNA polymerization (C) and after protein dissociation after completed polymerization (D). Different *E** values are expected for each FRET state, lower in the case of C and D states and unknown in the case of B.

In all experiments carried out in this Thesis, the Stoichiometry (*S*) should remain constant, relating the green and the red fluorescence after green excitation to the overall fluorescence emission after green and red excitation (Hohlbein *et al.,* 2014). In contrast, regarding FRET efficiency (*E**) we screened different DNA constructs to maximize the FRET change between the initial state before polymerization (medium FRET) to the final state after polymerization (low FRET). For that reason, endpoint assays for the DNA polymerization were performed. The assays were carried out

at room temperature in a test tube minimizing potential influences of very slow polymerization which cannot be monitored directly on the single-molecule level. Different DNA substrates (004-025, 004-026, 004-027, 004-028 and 004-040) were screened according to the protocol described in paragraph 2.2.27.

As shown (Fig. 3.25), the initial $E*$ values were different for each DNA construct. Modifying the donor positions by two bases supposed a difference of $E*$ roughly 0.05 units among the DNA substrates. The closer the donor dye was in respect to the acceptor dye, the higher the initial $E*$ was. Only in the case of 004-040 structure, in which the donor dye is placed on the free 5' end, no differences were found to sequence 004-025, indicating that the conformational flexibility of the single stranded overhang might provide similar distances despite using different labeling positions. After polymerization at room temperature using Klenow fragment under the above-mentioned conditions, different final $E*$ values were found. Here, the closer the donor dye was in respect to the acceptor dye, the lower the relative change in $E*$ upon polymerization (dE) was found. 004-040 resulted in the highest dE (from 0.53 to 0.19; dE=-0.34), followed by 004-025 (from 0.55 to 0.39; dE=-0.16) (Table 13). Only 004-028 DNA construct showed a higher $E*$ after polymerization (from 0.69 to 0.71; dE=0.02). Therefore, both 004-025 and 004-040 sequences were chosen for real-time binding and polymerization assays as they allow the highest contrast of $E*$ between the non-polymerized and the polymerized state. In addition, the assay was very reproducible when the 004-025 DNA substrate was subjected to endpoint assay using all 4 polymerases (Klenow fragment, HoLaMa, α subunit or the αεθ). A final $E*$ value close to 0.35 (Table 14) was obtained in all cases after full polymerization under the conditions already defined in section 2.2.27. Hence, these DNA constructs were suitable for successive real time polymerization experiments.

With regards to the binding experiments, in which the 004-025 and 004-040 DNAs were used, we observed changes in the transfer efficiency between donor and acceptor on our DNA constructs upon binding of various DNA polymerases. Nevertheless, the shift of $E*$ was different among polymerases and depending on the DNA substrate used for the binding experience. The initial $E*$ value of the DNA 004-025 was between 0.53 and 0.55 (Table 15), whereas initial $E*$ value of the 004-040 DNA substrate was equal to 0.55-0.57 (Table 16) (Supplementary information about protein binding in Appendix III).

Klenow fragment showed a shift of E^* to a higher level when bound to 004-025 (Fig. 3.27.A) and a shoulder at 0.78 when it bound to 004-040 (Fig. 3.28.A). On time traces, stable binding events were observed with a change in E^* after binding.

HoLaMa, whose structure consists of the polymerase domain from Klenow Fragment with 11 site-specific mutations that improve protein stability (paragraph 4.3), showed a different behavior upon binding to 004-025 DNA substrate than Klenow fragment, used as control. Binding of HoLaMa resulted in a shoulder at E^* equal to 0.71 in the FRET histogram (Fig. 3.27.B), representing the linked conformation of the DNA upon binding of the polymerase. This pattern was also seen for the 004-040 DNA construct: the shoulder, however, was centered around E^*=0.84 (Fig. 3.28.B). In both cases, binding events were stable but not as long as for Klenow fragment, indicating a higher dissociation constant of the HoLaMa-DNA complex. How does the lack of the 3' → 5' exonuclease domain can affect protein binding and why does the absence of this exonuclease domain lead to this increment of E^*, much more pronounced than in the case of Klenow fragment?

Beese and co-workers (1993b) determined that the DNA enters the polymerase active site of Klenow Fragment towards a cleft between the polymerase and exonuclease domains with an inclination around 80°. The absence of exonuclease domain in HoLaMa may lead to an increase in E^* upon binding much higher than KF: the donor dye might become closer to the acceptor probe because the absence of proofreading domain allows a reposition of the ssDNA with a higher inclination. In the case of the 004-040 DNA substrate, the donor dye is on the 5' terminus of the single-strand and, due to the higher inclination, may have access to the acceptor dye, thus increasing E^* quite more pronouncedly than in the case of the Klenow fragment. Moreover, the disruption of the binding site inevitably leads to an impairment of the polymerase attachment to the DNA substrate. The low polymerase activity of HoLaMa, as described in the following paragraphs, could be actually related to a reduced ability of the enzyme in promoting the assembly of the substrates into a catalytic proficient geometry.
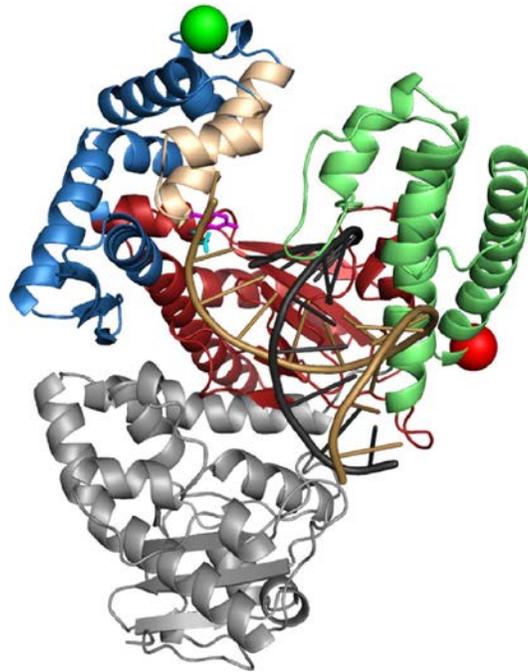
Figure 4.13: Open KF-DNA binary complex (PDB 1L3U).Fingers domain is coloured in blue (except O- and O1 in beige), Thumb domain in green, Palm domain in red and the 3' → 5' exonuclease domain in grey. Also a donor and an acceptor dye are present and represented by the green and the red spheres, respectively. The DNA template is shown in brown and the primer in black (Hohlbein *et al.,* 2013)

The binding affinity of HoLaMa was significantly lower than that of Klenow fragment when we used the one nucleotide gapped-DNA (003-025-032 DNA construct). The gap in this DNA construct was 1 nucleotide and the construct shows an increase in transfer efficiency when the polymerase binds (Fig. 4.14).
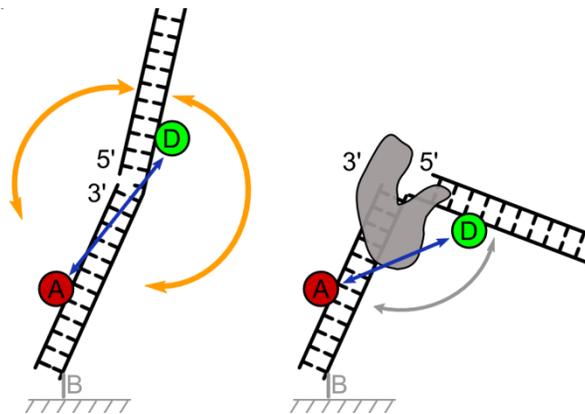
Figure 4.14: Scheme showing the changes in energy transfer on a 1 nucleotide-gapped DNA upon polymerase binding.

147

K_D constant was 70 pM for KF and 1.7 nM for HoLaMa. Due to this impairment of protein binding in the case of HoLaMa, we tested the effect of adding the complementary nucleotide to stabilize the binding. For this purpose, 100 µM dTTP was added to the imaging buffer. The 003-025-032 DNA substrate included a 3'-terminus dideoxi-terminated (C_H3') to block the polymerization process, restricting its use only for binding studies. The occupancy of DNA molecules to which HoLaMa bound improved from 27.4% to 81% when the complementary dTTP was present, compared with the fraction of the same DNA molecules to which HoLaMa bound in the absence of dTTP (Fig. 3.29.1). Furthermore, the binding of HoLaMa to the 003-025-032 binding sensor was more stable in the presence of complementary dTTP, represented by longer time-traces (Fig. 3.29.2).

Other authors have also shown that the presence of the complementary nucleotide stabilizes the DNA template-DNA polymerase complex (Luo *et al.,* 2007; Berezhna *et al.,* 2012; Markiewicz *et al.,* 2012). Moreover, those authors showed longer stable binding events when the complementary dNTP was present, stabilizing the tertiary structure formed by the DNA template, the nucleotide and the DNA polymerase in a Finger-closed conformation. Addition of DNA led the equilibrium towards an open state. The binding of a complementary nucleotide supposed the closing of the Fingers domain, whereas the ternary complex in the presence of an incorrect nucleotide shifted to a partially closed (*ajar*) conformation. Therefore, the process of Fingers closing is essential to guarantee a proper selection of the incoming nucleotide and a subsequent high fidelity during DNA replication.

Moreover, by using the ALEX technique and a double-labeled Klenow fragment (the donor dye at the Fingers domain and the acceptor dye at the Thumb domain) it was concluded that the inter-fluorophore distance between the open and the ajar conformation was 0.2 nm, whereas the distance is 1 nm between the open and closed conformations (Hohlbein *et al.,* 2013) (Fig. 4.15).
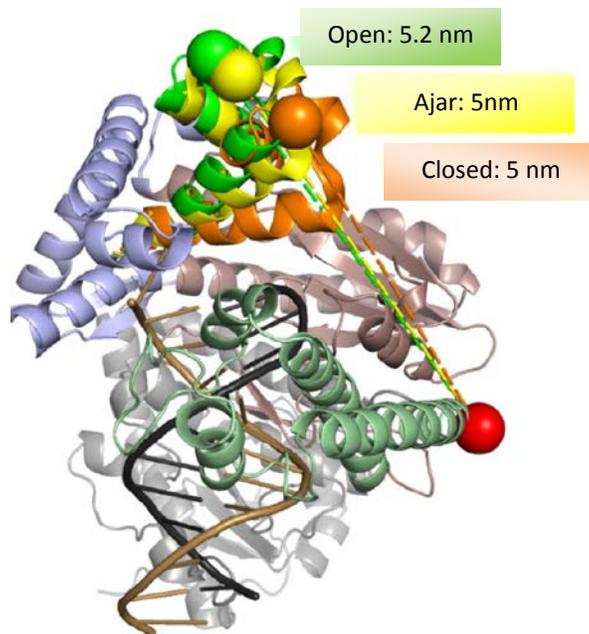
Figure 4.15: Polymerase conformation featuring an intermediate-FRET state. The crystal structure of the Pol-DNA binary complex (PDB 1L3U) is superimposed on the crystal structures of a ternary complex with a partially closed (*ajar*) conformation of the Fingers domain (PDB 3HPO) and another ternary complex with closed Fingers conformation (PDB 1LV5). The distances between the donor dye (green sphere) and the acceptor dye (red sphere) are highlighted in different colors, depending on the Fingers domain conformation, directly related to dNTP complementarity. (Hohlbein *et al.,* 2013).

The most intriguing situations were found with DNA Pol III. Regarding α subunit, there was not shift of energy transfer upon protein binding when we used the 004-025 DNA construct (Fig. 3.27.C). However, short trajectories of lower *E** were perceived on time traces (Fig. 3.27.C) and, therefore, there is a fair possibility that a transient binding and ulterior dissociation are present in the case of α subunit bound to this DNA substrate. On the contrary, two new unexpected peaks of *E**, one higher (0.72) and one lower (0.39) compared to the original peak (0.55) were observed on the histogram representing α subunit binding to the DNA construct 004-040 (Fig. 3.28.C). These differences could be linked to a free disposition of the single strand when we only consider the single α polymerase without the rest of the proteins co-involved in the catalytic core. This free disposition could lead to a movement of the single strand and its donor dye, located at the 5' terminus of the strand, approaching (higher *E**) or distancing (lower *E**) from the acceptor dye. Moreover, different high and low *E** frames were noted from the time traces (Fig. 3.28.C). For the time traces of both DNA constructions (004-025 and 004-040), a lower *E** was observed, whereas

a higher FRET level was detected only in the case of the 004-040 DNA (Fig. 3.28.2). On the one hand, the binding of α subunit could be represented by a lower E^*, since it was common in the traces of both DNA substrates; on the other hand, the observed traces with a lower E^* value could be related to a dissociation event and, in that case, binding would be observed only when the donor dye is at the 5'-terminus of the ssDNA. When our model was explained (Fig. 4.12), it was unknown whether and how E^* would change after the polymerase binds to the free 3' end of the DNA substrate, in contrast with DNA polymerization, which is supposed to be represented by a decrease in E^* as a result of an increase in the distance between the two dyes. Consequently, there is a possibility that α polymerase binding is related to a different E^* change when compared with KF, HoLaMa or the catalytic core.

Finally, αεθ binding resulted in a significant and unexpected shift of energy transfer. A shoulder at 0.96 and 0.88 appeared after the binding of αεθ to the 004-025 (Fig. 3.27.D) and 004-040 DNA substrate (Fig. 3.28.D), respectively. Time traces showed a long stable binding with a significant change of E^*. The most surprising aspect working with αεθ and these two DNA sensors is the fact that energy transfer curve shifted close to 1 after protein binding. Besides, the binding of HoLaMa to the 004-040 DNA construct also shifted up to a very high value close to 1 (Fig. 3.28.B). The E^* value has to be comprised between 0 and 1 ($0 \leq E \leq 1$) (Kapanidis *et al.,* 2005; Holden *et al.,* 2010), considering the equation already mentioned in paragraph 2.2.35,

$$E^*= DA/(DD+DA)$$

assuming values close to 0 for only-donor species (DD) and values below 1 for molecules with both donor and acceptor probes (DA). Neither α subunit nor Klenow Fragment showed this phenomenon after the attachment to the DNAs. Some structural aspects or mechanisms as DNA bending after protein binding could be taken into consideration. DNA bending has been observed in other DNA Pol III as that one from *Thermus aquaticus,* in which loops connecting the Thumb and Palm domains tends to bend DNA due to its high flexibility (Wing *et al.,* 2008). The absence of a complete crystal structure of *E. coli* αεθ of DNA Pol III makes it difficult to establish conclusions about that.

This remarkable rise in energy transfer, especially comparing with the increase regarding α subunit binding, could be related to the high flexibility of the Q-linker that connects the exonuclease subunit with the α subunit (Ozawa *et al.,* 2013) (Fig. 4.16). Several conformations were obtained for this 22-residues segment (Thr183-Thr201) that connects both subunits, showing its notable flexibility. In addition, it was found that the great mobility of the long Q-linker enables the ε subunit to approach the DNA and reorient itself to a suitable position. These conformational changes would lead the donor dye to approach the acceptor molecule, with the corresponding increase in energy transfer between dyes.
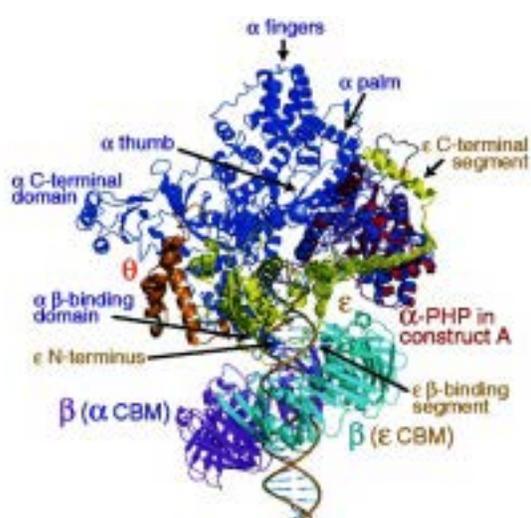
Figure 4.16: Models of the αεθ:β2 complex with primer-template DNA. The complex is presented in a closed conformation (Ozawa, *et al.,* 2013). α subunit is presented in dark blue, the PHP domain is shown in red and ε subunit is in yellow. Multiple rearrangements were observed for the Q-linker segment connecting α and ε subunits.

The polymerase activity of Klenow fragment, HoLaMa, α subunit and αεθ was determined in real time using those DNA substrates that presented the largest change in FRET efficiency from the initial state to the final one after polymerization in the endpoint assay (004-025 and 004-040 DNA substrates). These experiments were performed as described in section 2.2.33, with the purpose of analyzing changes of occupancy of different FRET species as a result of DNA polymerase activity, shown in histograms and time traces (Fig. 3.30 and 3.31; additional time-traces in Appendix IV).

E* of the 004-025 DNA substrate shifted from 0.54 to 0.35 (Table 18), whereas *E** shifted from 0.55 to 0.2 when 004-040 DNA substrate was used for evaluating polymerization in real time (Table 19). Figures 6 and 7 represent peaks of energy transfer before (top panel) and after three minutes (middle panel) and ten minutes (bottom panel) of polymerization of 004-025 and 004-040 DNAs, respectively, for all four polymerases.

Addition of Klenow fragment to the DNA construct resulted in a single peak in the energy transfer histogram centered at 0.35 (004-025 DNA) or 0.2 (004-040 DNA) after ten minutes of polymerization, indicating fully polymerized molecules. Time traces showed a slight increase of $E*$ upon protein binding immediately followed by a pronounced shift up to the expected value for full polymerization. The fact that polymerization took place immediately after protein binding indicated the high speed of reaction of this polymerase. Moreover, the average duration for all polymerization events observed on time traces was 2s (from 0.5 to 7.5 seconds) (Fig. 3.32.A), meaning that Klenow is able to incorporate 12 nucleotides per second. This value was slightly lower than other previous results observed for this polymerase (about 50 nucleotides per second, Olsen *et al.,* 2013). Slower traces of polymerization were observed when 1 µM dNTPs (Fig. 3.31.3.B), instead of 100 µM, was included in the imaging buffer. While with 100 µM dNTPs most of the traces showed a continuous shift from 0.78 (protein binding) to 0.2 (polymerized DNA), pausing events with constant $E*$ were visible at lower nucleotide concentrations (1 µM).

A different behavior was observed in the case of HoLaMa, due to the fact that a three peaked-histogram of energy transfer was reached in real time polymerization for HoLaMa polymerase. The peak at about 0.55 represents not polymerized species or those to which polymerase was not bound; the high-level FRET features a value of $E*$ equivalent to that represents protein binding (0.7 for 004-025 and 0.88 for 004-040 DNA substrate); the low-level FRET corresponds to polymerized molecules. From time traces, a slow reaction was observed, with long traces from the upper $E*$ value to the lower one (Fig. 3.30.2.B and 3.31.2.B), representing the low replication speed of the enzyme. The average duration of all polymerization events, observed on time traces, was estimated as 5 seconds (from 1s to 14s) (Fig. 3.32.B). Thus, HoLaMa can incorporate 5 nucleotides per second.

However, the expected $E*$ after complete polymerization of the DNA substrate 004-040 (0.2) was not reached within ten minutes after adding polymerases and nucleotides within the imaging buffer to the immobilized DNA, probably due to the weak initial binding of HoLaMa to the DNA. Most likely, a longer period of polymerization must be evaluated for obtaining a value equal to 0.2 (as in the case of Klenow fragment or DNA Pol III). Unfortunately, photobleaching (destruction of dye fluorescence by the effect of light and/or molecular oxygen) did not allow us to measure for longer periods.

An advantageous alternative to maximise observation time could be the use of magnetic tweezers, where a small magnetic bead applies stretching forces to an immobilised DNA molecule, enabling the visualization of conformational changes without exerting photobleaching of fluorophores (Long *et al.,* 2013; Funabashi *et al.,* 2014), although probably the method is not as sensitive as the one we used. Also the use of novel fluorophores or new oxygen scavengers is being developed (Zheng *et al.,* 2014) to reduce photobleaching.
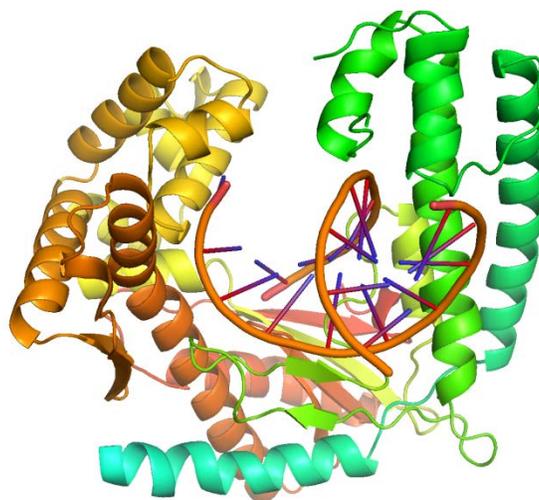
Figure 4.17: HoLaMa complexed with mismatched double-stranded DNA. Protein structure was obtained from Klenow Fragment structure (PDB 2KZZ), removing the 3' → 5' exonuclease domain. DNA was obtained from PDB file 1QTM.

As discussed in paragraph 4.3, the low polymerization speed of HoLaMa may be due to the absence of the 3' → 5' exonuclease domain, which is responsible for controlling the DNA to come into the polymerase active site (Beese *et al.,* 1993b). Therefore, despite not being directly involved in polymerase activity, the 3' → 5' exonuclease domain of Klenow Fragment seems to be important for the correct development of DNA elongation, impairing kinetics in HoLaMa.

The α subunit showed FRET characteristics similar to that of Klenow fragment. The initial peak representing the DNA substrate molecules (0.53 for 004-025 DNA and 0.55 for 004-040) shifted to *E** values equal to 0.31 and 0.22 for the 004-025 (Fig. 3.30.1.C) and the 004-040 (Fig. 3.31.1.C) DNA substrates, respectively. The initial population disappeared completely after ten minutes of

polymerization. The high speed of polymerization of α subunit was also observed on time traces (Fig. 3.30.2.C and Fig. 3.31.2.C). Most of the molecules showed an average polymerization duration of around 1 second (from 0.5 to 5 seconds) (Fig. 3.32.C). Consequently, an incorporation rate of 25 nucleotides per second was estimated for α subunit alone, which is twice the rate of KF under the same conditions (50 pM enzyme, 100 μM dNTPs). The α subunit associated to the sliding clamp, however, was previously determined to be 1000 bp/s (Maki *et al.,* 1988), highlighting the role of the clamp as an important processivity factor. Nevertheless, our result is comparable to those results for α subunit alone determined by Fay and co-workers (1982), who measure a polymerization rate of 10 nucleotides per second.

Finally, regarding the αεθ, histograms showed the presence of three peaks of *E**: 0.33, 0.57 and 0.95 for the 004-025 DNA substrate (Fig. 3.30.1.D); 0.20, 0.56 and 0.9 for the DNA 004-040 (Fig. 3.31.1.D). As in the case of HoLaMa, in which a three-peaked histogram was also present, the first peak represents the occupancy of polymerized molecule species since the peak reached the expected value according to the endpoint assay; the medium-FRET level corresponds to the molecules not yet polymerized and the high-FRET level occupancy is related to those molecules to which αεθ is bound. As shown in the histograms, the peak representing full-polymerized increased upon 10 minutes of polymerization. The occupancy of these polymerized molecules was 21% and 57% after 3 and 10 minutes of polymerization, respectively (16% and 48% in the case of α subunit). The mean duration of polymerization events in most of the molecules was also shorter than in the case of α subunit: about 0.5 seconds (ranging from 0.5 to 3 seconds) (Fig. 3.32.D). Therefore, the αεθ catalyzed the reaction faster than the single α subunit after three or ten minutes of polymerization, respectively, and under the same conditions (50 pM enzyme, 100 μM dNTPs). Hence, it is the first time that the stimulating effect of ε subunit on polymerase activity of *E. coli* DNA Pol III α subunit is visualized at the single molecule level.

In the case of αεθ, the presence of other co-involved enzymatic activities, as the exonuclease activity of ε subunit, must be further discussed. The switching from pol-site to of α subunit to the exo-site of ε subunit is a direct consequence of the misincorporation of an incorrect dNTP (Tsoi *et al.,* 2003; Markiewicz *et al.,* 2012). However, there are no data about conformational changes at the single-molecule level after DNA switching to the exo-site of ε subunit during DNA replication by *E. coli* DNA Pol III. The highly mobile Q-linker (Fig. 4.16) allows ε to approach the DNA with the

corresponding conformational rearrangements (Ozawa *et al.,* 2013) that could be involved in energy transfer modification, as it was previously mentioned regarding protein binding. Only in the presence of mismatched nucleotides at the primer terminus, which can be misincorporated during DNA elongation, the DNA template will bind the exo-site (Tsoi *et al.,* 2003; Markiewicz *et al.,* 2012). Christian and co-workers (2009) observed dynamics related to the movement of the primer from the pol-site to the exo-site of DNA Pol I after the incorporation of an incorrect nucleotide.

Moreover, the switching between both sites in DNA Pol I has been defined by two different mechanisms: dissociating from the enzyme (intermolecular transfer) or remaining closely attached to the polymerase (intramolecular transfer) (Lamichhane *et al.,* 2013) (Fig. 4.18).
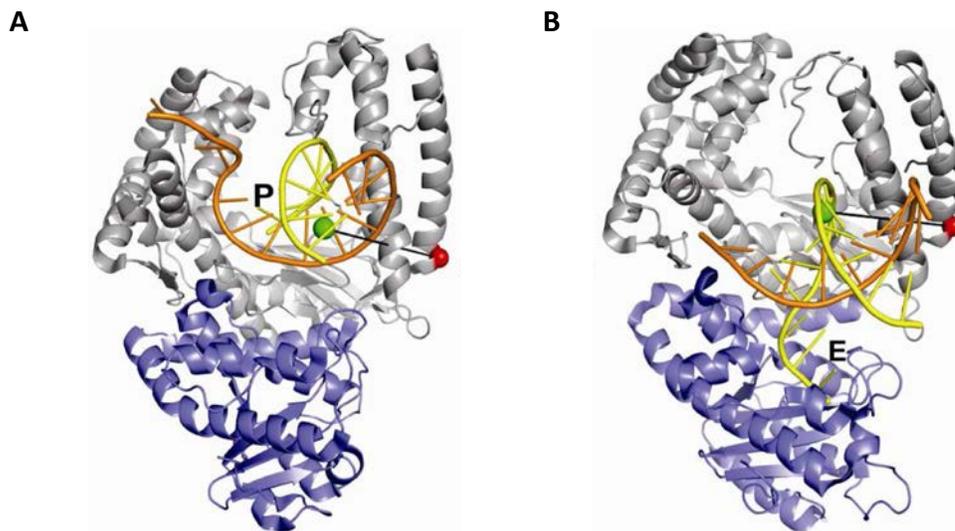
Figure 4.18: DNA switching from the pol-site (A) to the exo-site (B) of Klenow fragment. DNA is labeled with donor dye (green sphere), whereas the acceptor dye (red sphere) is located on the exonuclease domain (Lamichhane *et al.,* 2013).

After the cleavage of the misincorporated dNTP at the primer terminus, a new switching from the exo-site to pol-site is also required (Fidalgo da Silva & Reha-Krantz, 2007) in order to continue DNA elongation. However, the study of the switching between pol- and exo-sites has not been carried out with DNA Pol III α and ε subunits at the single-molecule level. This switching between two different subunits may influence the efficiency of energy transfer during DNA elongation since

DNA unwinding is necessary before performing the proofreading activity (Cowart *et al.,* 1989) and we are working with a double-labeled DNA.

In summary, we have developed a suitable protocol for the immobilization of double labeled-DNA molecules by a PEG-neutravidin system to monitor DNA synthesis in real time. DNA molecules are subjected to laser excitation in order to analyze the distance-dependent energy transfer between donor and acceptor dyes by smFRET. The changes in FRET enable us to determine different species and characterize conformational changes upon protein binding and DNA polymerization. The four assayed DNA polymerases catalyzed the elongation of the DNA strand with the subsequent decrease in FRET between donor and acceptor dyes since the distance between them was increased upon polymerization. By comparing the different polymerases we were able to characterize the speed of replication together with structural rearrangements of the ssDNA. Nevertheless, further studies are required to incorporate new information about the complex machinery of polymerization systems present in microbiology at the single molecule level. New crystal structures of complete DNA Pol III would be a perfect tool to understand some protein behavior observed at the single-molecule level.

CONCLUSIONS AND FUTURE PERSPECTIVES

The present Thesis has highlighted, for the first time, evidence about the *in vivo* assembly of the DNA Pol III of *E. coli*. By co-expressing DNA Pol III α, τ, ε and θ subunits in a dual-plasmid co-expression system, a trimeric ατεθ replicase was obtained. A molecular model for this trimeric replicase has been defined, after detecting polymerase-polymerase contacts. In addition, kinetic studies demonstrated how two of its α subunits are responsible for the extension of the double-primed DNA (mimicking the lagging strand) whereas the remaining α subunit catalyzes the replication of the single-primed DNA (simulating the leading strand). The replication of the double-primed DNA requires an inter-primer distance of at least 40 nucleotides. Moreover, the extension of this DNA by two concomitant α subunits slows down the replication of the single-primed DNA, pointing out the coordination of the synthesis of the two DNA strands. Considering the success of the dual-plasmid system for protein co-expression, new alternatives appear for the complete expression of the DNA Pol III holoenzyme. At least 4 plasmids should be used for a suitable protein yield. Combined plasmids must guarantee the corresponding compatibility of their origins of replication, antibiotic resistances and induction systems, leading to the co-expression of the complete clamp loader together with DNA Pol III*. The genes encoding the γ, δ, δ', ψ and χ subunits would be cloned into a third new low-copy number plasmid, presumably yielding a good protein expression level; finally, the inclusion of the sliding clamp encoding gene (*dnaN*), in a fourth high-copy number plasmid, would enable the entire overproduction of the DNA Pol III holoenzyme. In addition, attempts to the crystallization of DNA Pol III must be developed for completing the structural information about the only essential polymerase in *E. coli.*

In the present work, the PHP domain of *E. coli* α subunit was confirmed to catalyze a pyrophosphatase reaction, according to a type II pyrophosphatase activity, since it depends on Mn^{2+} as metal activator. H12 and D19 are responsible for developing the catalytic activity, due to the fact that their substitution with alanine strongly lowered the pyrophosphatase reaction. On the other hand, D201 controls the release of phosphate molecule at the interface with the solvent. The modification of the coupling between polymerase and pyrophosphatase activities after residue substitution was demonstrated, indicating for the first time the major role of the PHP domain on polymerase reaction, despite not being directly involved in DNA elongation. Actually,

the accumulation of inorganic pyrophosphate acts as product inhibitor of DNA replication. Our results provide information about the catalytic site of the PHP domain, which was also confirmed to lack exonuclease activity; it would be therefore of interest to complete the characterization of the active site. Moreover, because of the observed phenotypic differences after residue substitution, the PHP domain may become a new potential target for novel inhibitory molecules that could affect to the growth of Gram negatives.

A third part of the Thesis concerns HoLaMa, a mini-DNA polymerase designed by molecular modelling form the tertiary structure of the Klenow polymerase, whose 3' → 5' exonuclease domain was removed, whereas the polymerase domain was subjected to 11 site-specific mutations. This artificial polymerase showed good results of protein stability and solubility but lower polymerization speed, in contrast with Klenow enzyme. The lack of the proofreading domain was directly related to the slow DNA polymerase activity and, according to our observations, the exonuclease domain may confer processivity to the DNA polymerase activity. Moreover, Klenow enzyme is assumed to be a high-fidelity replicative enzyme; if HoLaMa was able to introduce a discrete number of random mutations during the replication process, it could be further engineered and used, for instance, in mutagenic PCR experiments. Thus, a comparison of the frequency of mutations during DNA replication between HoLaMa and Klenow enzyme may well be highly important for the success of this artificial enzyme. Considering the molecular mass of HoLaMa, NMR analysis could be a suitable approach for completing the information about polymerase conformation. Some mechanisms for increasing HoLaMa processivity might be further investigated with the purpose of achieving novel possibilities with this artificial polymerase.

Finally, the assay developed for the characterization of different polymerases at the single-molecule level by FRET enabled us to analyse polymerase binding and DNA polymerization by measuring changes in FRET efficiency. The assay was very reproducible, even working with different DNA substrates. Results obtained by this biophysical analysis were close to the results obtained by coupled-enzyme assays. The stimulating effect of ε subunit on polymerase activity was observed at the single-molecule level by the first time; furthermore, the improvement in protein binding in the presence of the complementary nucleotide was also detected. For a deeper analysis of DNA polymerases at the single-molecule level, new techniques must be developed. By using Cy3 donor dye and single-molecule Protein Induced Fluorescence Enhancement (PIFE) the

end of the polymerase reaction would be confirmed, since there is a change in stoichiometry when the polymerase contacts the Cy3 donor present at the 5'-terminus of the single-stranded DNA (Hwang *et al.,* 2011; Markiewicz *et al.,* 2012). The use of dark quenchers is another strategy to increase the concentration of the acceptor dye up to 1 µM minimizing the background. Dark quenchers are chromophores that relax from the higher electronic state to the ground state nonradiatively (Le Reste *et al.,* 2012), enabling the identification of the exact number of incorporated nucleotides and determining polymerization speeds. Lastly, the effect of the misincorporated dNTPs and the switching to the proofreading ε subunit must be analysed by FRET, due to the lack of data about *E. coli* DNA Pol III.

ACKNOWLEDGEMENTS

This thesis is the result of a long period of research. Time when many challenges, efforts, ideas, successes and not so good results have intermingled culminating in this work, which leaves me with a satisfactory feeling. However, all this work is not exclusively mine. Therefore, I wish to thank all those who in this period helped me, in one way or another, to achieve the challenge of completing my doctoral thesis.

It would be unfair not to start mentioning who allowed me to enjoy the unbeatable experience of doing a PhD in a foreign university: Cardinal Gil de Albornoz. His generosity allowed, more than 650 years ago, the foundation of one of the most prestigious, generous institutions ever created. The Royal Spanish College is nowadays synonymous of academic excellence, and it has also given me and hundreds of other students in the last six and a half centuries the opportunity to live the most important personal and professional experience. Thus, thanks to the Cardinal, but also to the Board of Trustees of the Royal Spanish College, for giving me this scholarship that has been a turning point in my personal and professional training. Remembering the Royal Spanish College will also bring to my mind indelible moments in which who has been our Rector, José Guillermo García-Valdecasas y Andrada de Vanderwilde, will always be the main protagonist. I will always remember his vocation for the College and his unconditional support for all of us.

Of course, I cannot forget to thank all of my college fellows, some of them have become true friends I am rather sure they will remain so for life. Thanks to Juan Pablo Murga and Salvador Tomás, for their honest friendship, loyalty and exemplary values; thanks to Manuel Parada, for his work as a "*godfather*"; of course, to José Torregrosa, for his trust in me and every moment we have enjoyed together; thanks to Enrique Sánchez de Castro, for his absolute demonstration of friendship; and finally, thanks to Enrique Gandía, Gustavo Díaz and José Manuel Macarro, for giving me the honor of sharing all our college time, from our first San Clemente's day to the important moment of our PhD defense.

The work described in this thesis derives from a research conducted at the Department of Pharmacy and Biotechnology of the University of Bologna. During this period, I had the fortune of being involved in a wonderful research team, directed by my supervisor Professor Alejandro

160

MASS SPECTROMETRY DATA

Swiss Prot cod DPO3A_ECOLI	Mascot score 8105	Unique peptides 57			
Peptide	Mr (expt)	Mr (calc)	Peptide	Mr (expt)	Mr (calc)
11-23	1434.6800	1434.6813	577-595	2004.0877	2004.0891
24-52	3064.6105	3064.6106	579-594	1676.8540	1676.8621
108-118	1088.5604	1088.5614	579-595	1804.9555	1804.9570
146-165	2355.0071	2355.0114	596-602	895.4218	895.4222
166-172	952.5382	952.5382	603-614	1366.6730	1366.6729
173-192	2213.0716	2213.0712	615-621	803.4576	803.4575
193-203	1139.6266	1139.6299	615-622	975.5542	975.5535
204-216	1550.6998	1550.7001	716-722	831.4154	831.4160
233-241	1185.5216	1185.5237	734-743	1075.5334	1075.5332
273-294	2549.2026	2549.2036	744-750	862.4798	862.4800
273-295	2721.2986	2721.2996	751-758	902.4278	902.4286
296-303	930.4786	930.4770	759-775	1937.0041	1937.0047
304-314	1364.6604	1364.6612	776-796	2342.0022	2342.0017
304-316	1605.8401	1605.8402	797-810	1645.8391	1645.8320
317-324	1119.5436	1119.5421	811-839	3165.6133	3165.6135
318-324	963.4392	963.4410	840-853	1409.7518	1409.7514
353-362	966.4884	966.4883	854-860	840.4228	840.4242
363-375	1192.6454	1192.6452	882-890	1024.5016	1024.5012
376-390	1834.9350	1834.9353	896-907	1218.6286	1218.6278
411-424	1687.8102	1687.8100	955-971	1959.0476	1959.0465
412-424	1531.7092	1531.7089	982-990	1125.5598	1125.5601
425-439	1567.7917	1567.7916	984-990	884.3806	884.3810
440-447	884.5174	884.5192	991-1004	1324.8181	1324.8191
448-458	1258.6456	1258.6459	993-1004	1139.7032	1139.7026
462-473	1267.6848	1267.6846	1047-1064	1880.9878	1880.9884
474-492	2206.0312	2206.0317	1069-1077	1092.4758	1092.4757
474-499	2976.4447	2976.4426	1083-1091	944.5294	944.5291
493-499	804.4164	804.4164	1092-1100	1113.5764	1113.5778
500-506	801.4708	801.4708	1101-1109	1134.6276	1134.6258
511-521	1048.6020	1048.6029	1103-1109	865.4406	865.4406
536-543	970.4856	970.4872	1110-1121	1432.7468	1432.7463
554-560	866.4654	866.4650	1128-1135	1005.5506	1005.5508
561-574	1658.9082	1658.9065	1148-1160	1434.6876	1434.6878

Table 23: Mass spectrometry data obtained with *in gel* digested proteins leading to the identification of *E. coli* DNA polymerase III α subunit. Sequence coverage 59%.

Swiss Prot cod	Mascot score	Unique peptides			
DPO3X_ECOLI	10326	39			
Peptide	Mr (expt)	Mr (calc)	Peptide	Mr (expt)	Mr (calc)
37-47	1300.6648	1300.6676	356-362	810.4502	810.4500
87-100	1576.8426	1576.8461	356-372	1956.0405	1956.0403
99-105	847.4388	847.4399	363-372	1179.5960	1179.5958
99-117	2203.1165	2203.1233	373-416	4550.3491	4550.3479
101-117	1973.9791	1973.9807	430-438	929.4936	929.4930
106-117	1373.6886	1373.6939	431-438	801.3980	801.3981
122-133	1473.7645	1473.7650	439-449	1209.6578	1209.6578
134-141	928.5130	928.5130	478-488	1316.6902	1316.6911
142-151	1177.5980	1177.5979	499-512	1563.8574	1563.8620
152-169	2012.1682	2012.1670	500-512	1435.7672	1435.7670
177-184	942.5122	942.5134	513-520	871.4760	871.4763
185-201	2101.0456	2101.0453	535-545	1269.7080	1269.7081
264-274	1157.6228	1157.6226	572-580	974.5466	974.5470
275-291	1978.0663	1978.0710	581-597	1813.9056	1813.9058
292-312	2183.1442	2183.1442	598-603	800.4174	800.4181
319-336	2104.1566	2104.1568	604-610	879.4334	879.4338
337-345	1087.5652	1087.5662	616-628	1485.7782	1485.7787
338-345	959.4708	959.4712	616-629	1641.8788	1641.8798
346-355	1204.4170	1204.6420	629-643	1835.9072	1835.9054
347-355	1048.5410	1048.5409	630-643	1679.8064	1679.8042

Table 24: Mass spectrometry data obtained with *in gel* digested proteins leading to the identification of *E. coli* DNA polymerase III τ subunit. Sequence coverage 62.8%.

APPENDIX I

MASS SPECTROMETRY DATA

Swiss Prot cod	Mascot score	Unique peptides
DPO3E_ECOLI	562	9
Peptide	**Mr (expt)**	**Mr (calc)**
2-7	807.2914	807.2929
8-29	2441.1629	2441.1645
30-41	1310.7568	1310.7558
30-42	1466.8564	1466.8569
127-135	930.5126	930.5134
143-151	1019.4702	1019.4706
152-159	1023.4984	1023.4985
160-204	4903.2799	4903.2868
214-228	1656.8102	1656.8107

Table 25: Mass spectrometry data obtained with *in gel* digested proteins leading to the identification of *E. coli* polymerase III ε subunit. Sequence coverage 56.4%.

Swiss Prot cod	Mascot score	Unique peptides
HOLE_ECOLI	240	5
Peptide	**Mr (expt)**	**Mr (calc)**
8-28	2234.1286	2234.1253
31-42	1390.6898	1390.6915
31-49	2280.1318	2280.1321
43-49	907.4528	907.4512
61-68	858.4920	858.4923

Table 26: Mass spectrometry data obtained with *in gel* digested proteins leading to the identification of *E. coli* DNA polymerase III θ subunit. Sequence coverage 63.1%.

STEADY-STATE ASSAYS OF α SUBUNIT
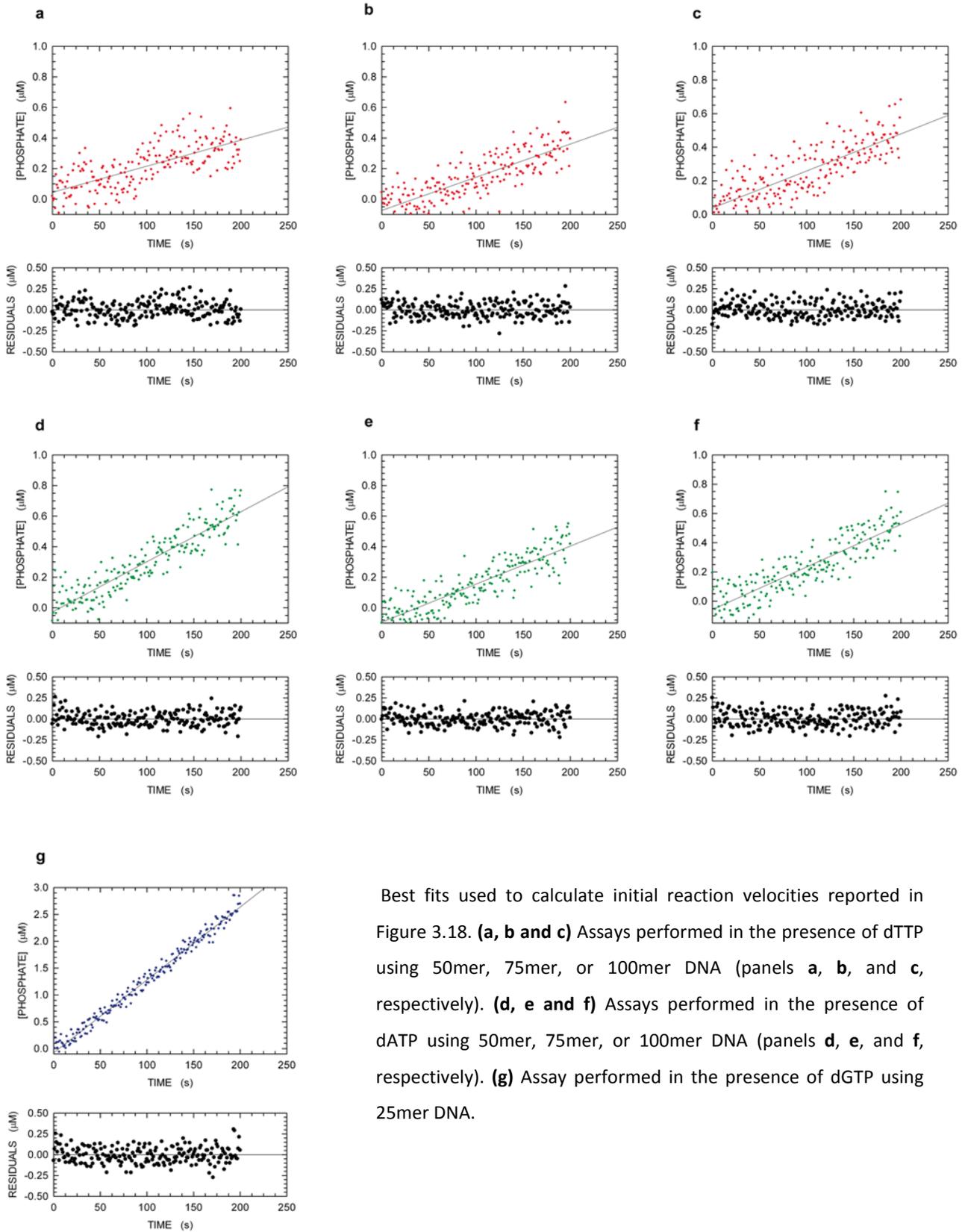
Best fits used to calculate initial reaction velocities reported in Figure 3.18. **(a, b and c)** Assays performed in the presence of dTTP using 50mer, 75mer, or 100mer DNA (panels **a**, **b**, and **c**, respectively). **(d, e and f)** Assays performed in the presence of dATP using 50mer, 75mer, or 100mer DNA (panels **d**, **e**, and **f**, respectively). **(g)** Assay performed in the presence of dGTP using 25mer DNA.

A
B
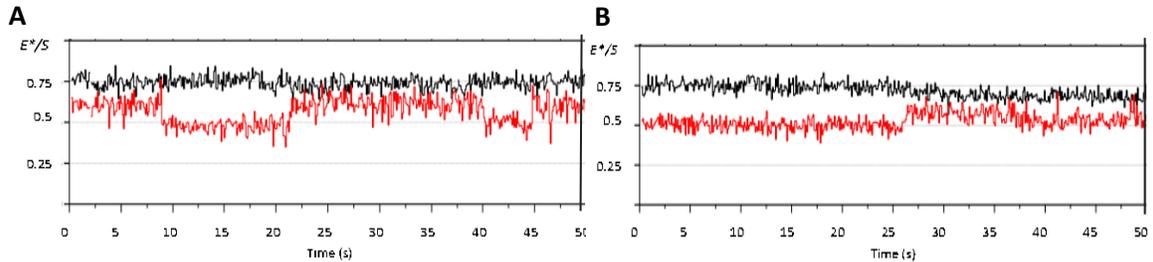
Exemplary FRET time traces representing binding of 2 nM of Klenow Fragment to the 004-025 DNA construct. Binding leads to a small change in *E**. **(A)** Binding for about 8 s followed by protein dissociation and a re-binding for another 19 s in a second event and for another 5 seconds in a third binding event; **(B)** Binding for 23 s. Black line represents stoichiometry; red line represents *E**.

A
B
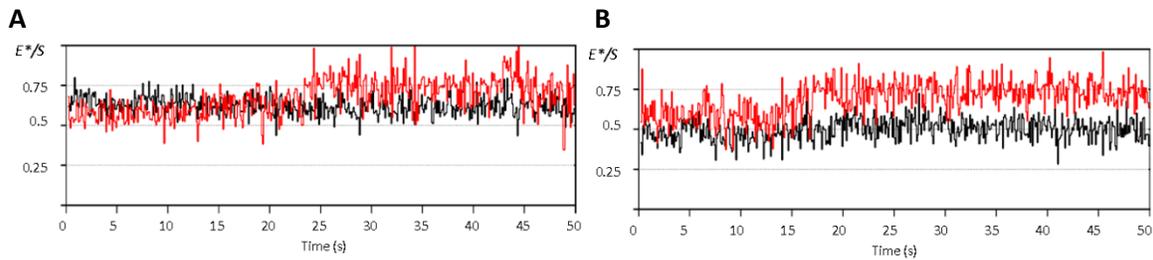
Exemplary FRET time traces representing binding of 2 nM of HoLaMa to the 004-025 DNA construct. Binding leads to a change in *E** to 0.70-0.75. **(A)** Binding for about 25 s; **(B)** Binding for 33 s. Black line represents stoichiometry; red line represents *E**.

A
B

C
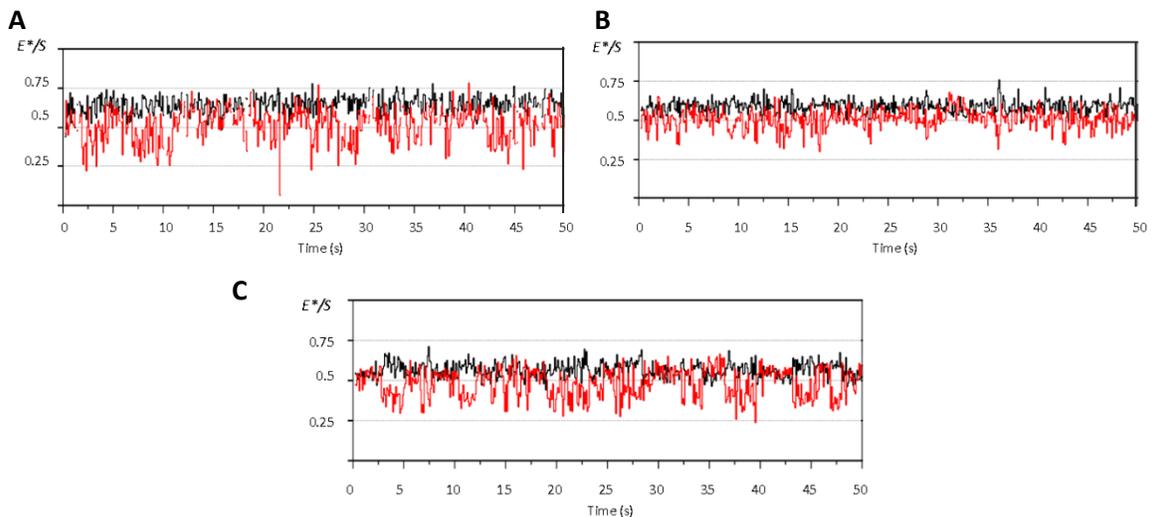
Exemplary FRET time traces representing binding of 2 nM of α subunit to the 004-025 DNA construct. Frames are representing binding and dissociation. Two FRET levels are detected. Black line represents stoichiometry; red line represents *E**.

TIME TRACES OF PROTEIN BINDING

A

B

Exemplary FRET time traces representing binding of 2 nM αεθ to the 004-025 DNA construct. Binding leads to a change in *E** to 0.95. Binding for 37 s (**A**) or 25 s (**B**). Black line represents stoichiometry; red line represents *E**.

A

B
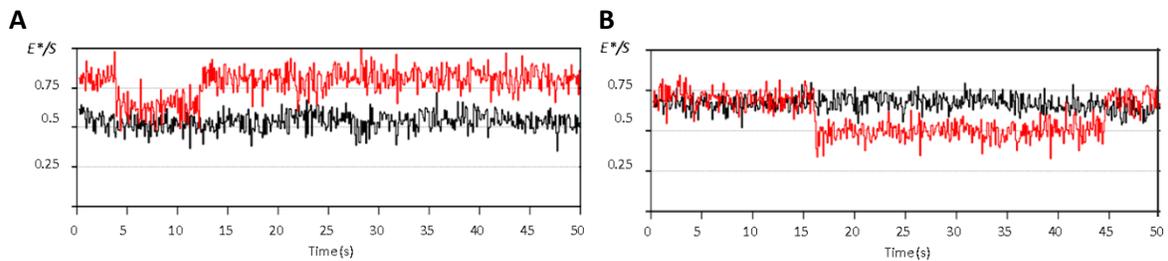
Exemplary FRET time traces representing binding of 2 nM of Klenow Fragment to the 004-040 DNA construct. Binding leads to a change in *E** to 0.75-0.8. **(A)** Binding for about 4 s followed by protein dissociation and a re-binding for another 33 s; **(B)** Binding for 16 s followed by protein dissociation and a re-binding for another 5 s. Black line represents stoichiometry; red line represents *E**.
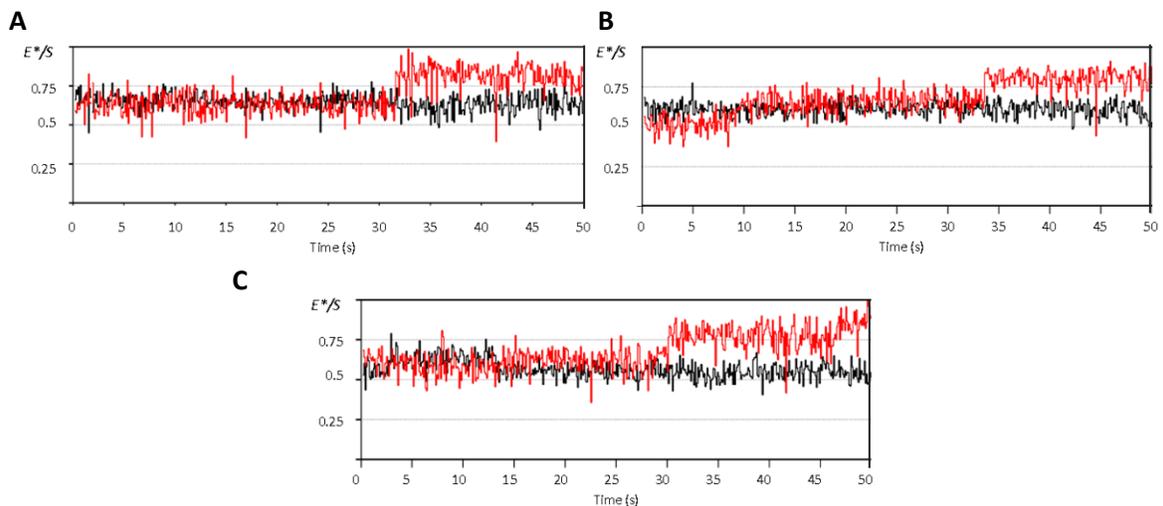
A

B

C

FRET time traces representing binding of 2 nM of HoLaMa to the 004-040 DNA construct. Binding leads to a change in *E** to 0.85-0.9. Binding for about 17 s (**A**), 15 s (**B**) or 20 s (**C**). Black line represents stoichiometry; red line represents *E**.

APPENDIX III

TIME TRACES OF PROTEIN BINDING
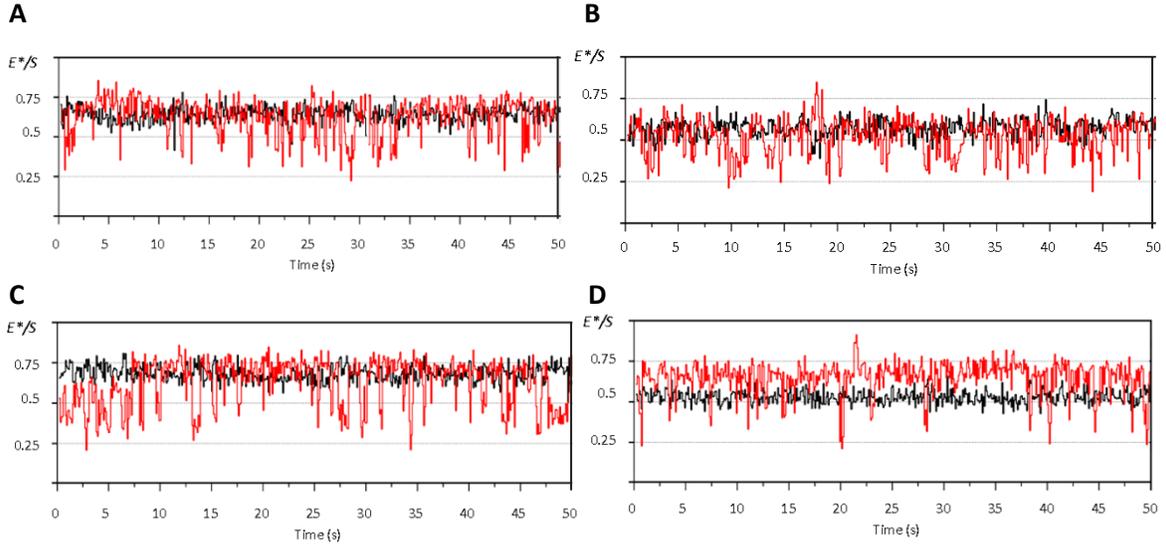
Exemplary FRET time traces representing binding of 2 nM of α subunit to the 004-040 DNA construct 2 nM of α subunit. Frames represent binding and dissociation. Three FRET levels (low, middle and high) are detected because of the movement of the ssDNA overhang upon protein binding. As discussed in chapter 4.4, is unclear how protein binding is represented in time-traces. Black line represents stoichiometry; red line represents E^*.
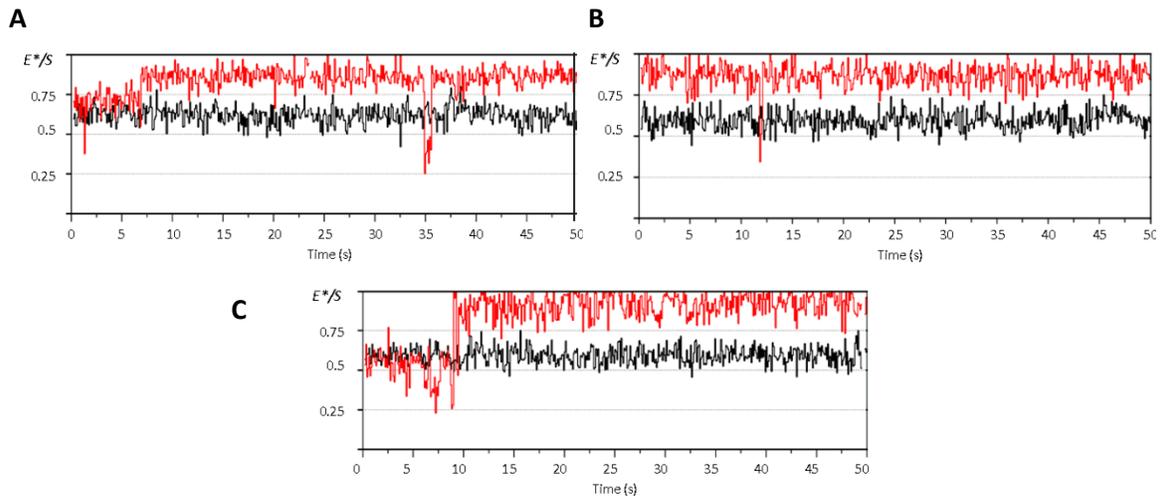
Exemplary FRET time traces representing binding of 2 nM αεθ to the 004-040 DNA construct. Binding leads to a change in E^* to 0.9. **(A)** Binding for 28 s followed by dissociation of 1.5 s and new binding event for 14 s; **(B)** Binding for 6 s followed by a rapid dissociation and new binding event for 38 s; **(C)** Binding for 40 s. Black line represents stoichiometry; red line represents E^*.

APPENDIX IV

TIME TRACES OF REAL TIME POLYMERIZATION

A

B
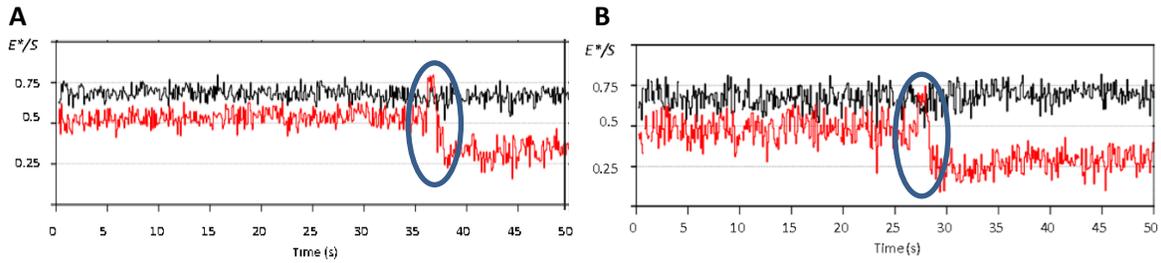
Exemplary FRET time traces representing polymerization in real time of the 004-025 DNA construct in the presence of 100 μM of each dNTP and 500 pM of KF. Both traces feature polymerase binding for around 2-3s with a short increase of FRET efficiency, followed by a decrease in *E** to 0.35, representing polymerization. Black line represents stoichiometry; red line represents *E**. The blue ellipses highlight the polymerization events.

A

B

C

D

E

F
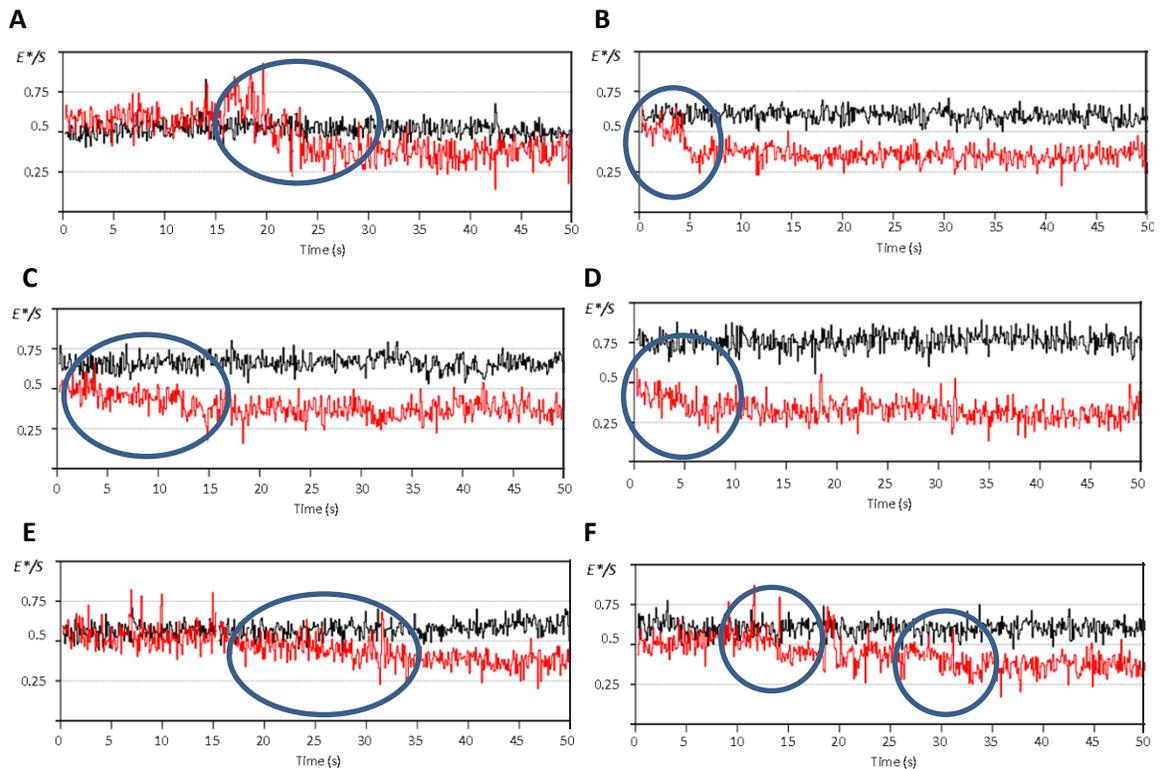
Exemplary FRET time traces representing polymerization in real time of the 004-025 DNA construct in the presence of 100 μM of each dNTP and 2 nM of HoLaMa. Polymerization (highlighted by the blue ellipses) is represented by a decrease in *E** to 0.35 in about 13 s (**A**), 4 s (**B**), 13 s (**C**), 8 s (**D**), 18 s (**E**) and two events of 8 s each (**F**). Black line represents stoichiometry; red line represents *E**.
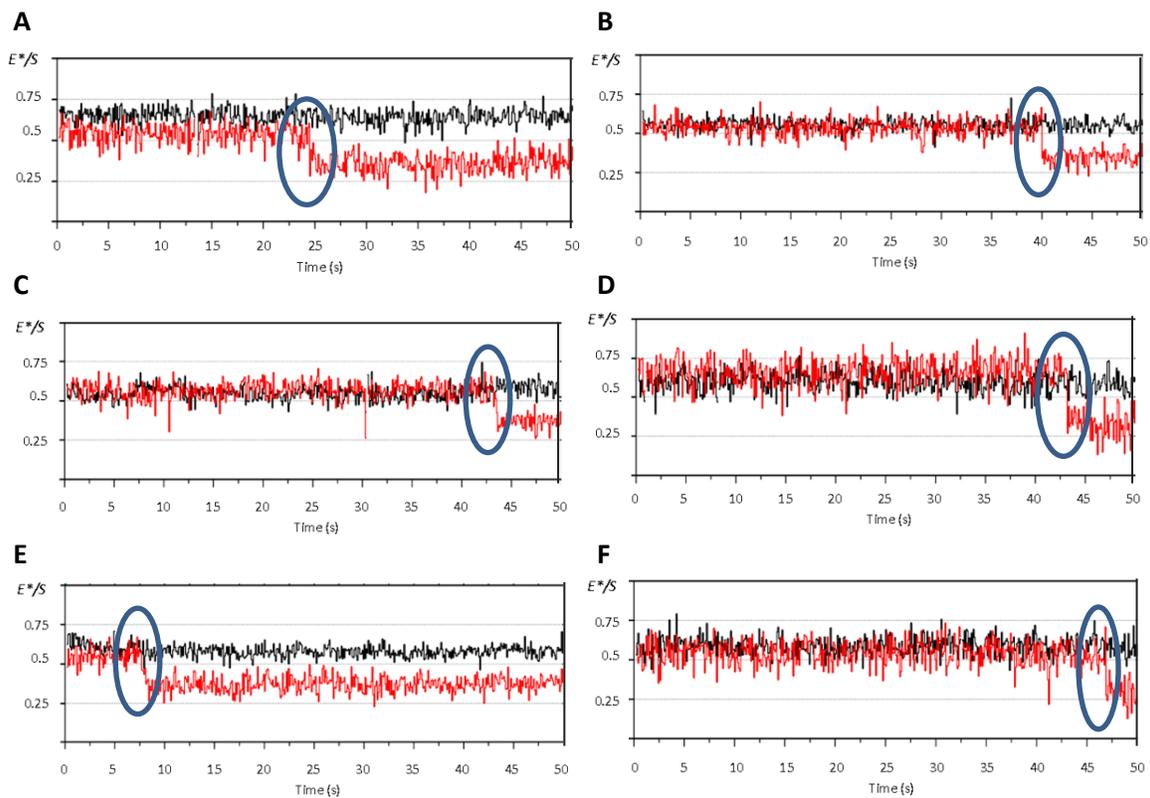
Exemplary FRET time traces representing polymerization in real time of the 004-025 DNA construct in the presence of 100 µM of each dNTP and 50 pM of α subunit. Polymerization (highlighted by the blue ellipses) is represented by a decrease in *E** to 0.35 in about 2.5 s (**A**), 0.5 s (**B**), 1.5 s (**C**), 1.5 s (**D**), 2 s (**E**) and 1.5 s (**F**). Black line represents stoichiometry; red line represents *E**.
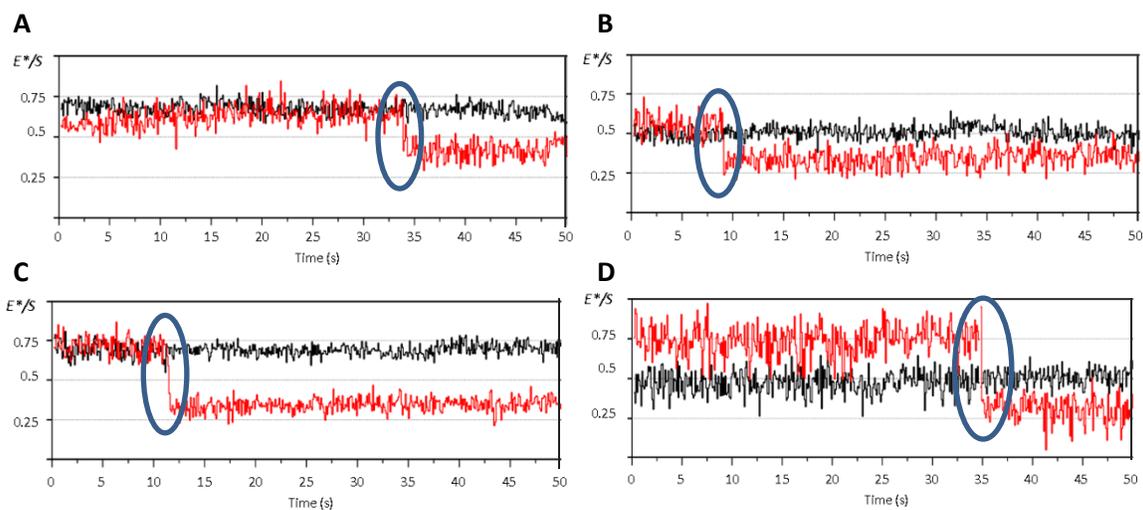
FRET time traces representing polymerization in real time of the 004-025 DNA construct in the presence of 100 µM of each dNTP and 50 pM of αεθ. Polymerization (highlighted by the blue ellipses) is represented by the decrease in *E** to 0.35 in about 0.5 s in all traces. Black line represents S; red line represents *E**.

APPENDIX IV

TIME TRACES OF REAL TIME POLYMERIZATION
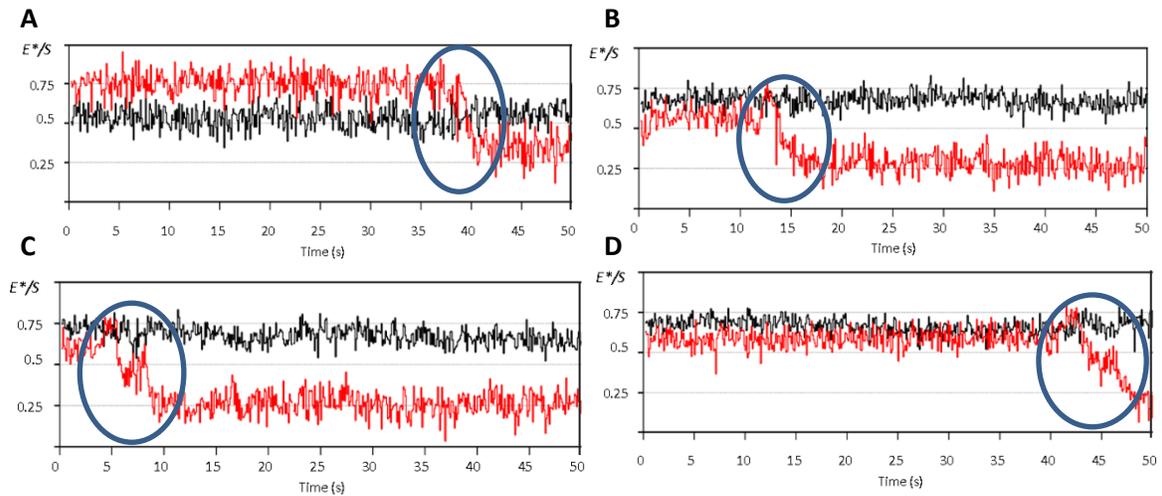
Exemplary FRET time traces representing polymerization in real time of the 004-040 DNA construct in the presence of 1 µM of each dNTP and 50 pM of KF. Traces feature polymerase binding with a short increase of FRET efficiency, followed by a decrease in $E*$ to 0.2, representing polymerization, in about 5 s (**A**), 6 s (**B**), two events of about 2 s each (**C**) and two events of 2.5 and 3 s (**D**). Black line represents stoichiometry; red line represents $E*$. The blue ellipses highlight the polymerization events.
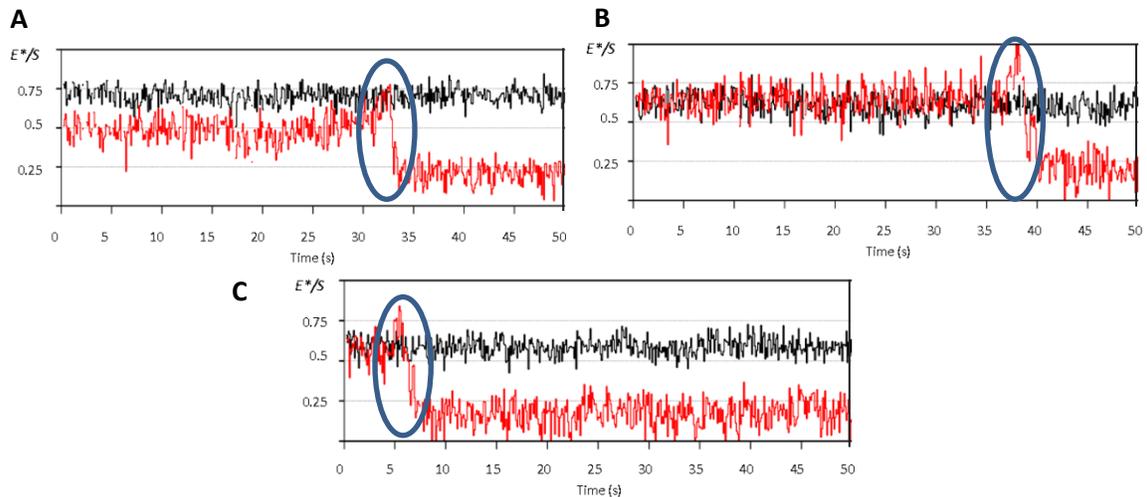
Exemplary FRET time traces representing polymerization in real time of the 004-040 DNA construct in the presence of 100 µM of each dNTP and 50 pM of KF. Traces show polymerase binding with a short increase of FRET efficiency (2 s) followed by a decrease in $E*$ to 0.2, representing polymerization, in about 2 s (**A**), 3 s (**B**), and 2.5 s (**C**). Black line represents stoichiometry; red line represents $E*$. The blue ellipses highlight the polymerization events.

TIME TRACES OF REAL TIME POLYMERIZATION
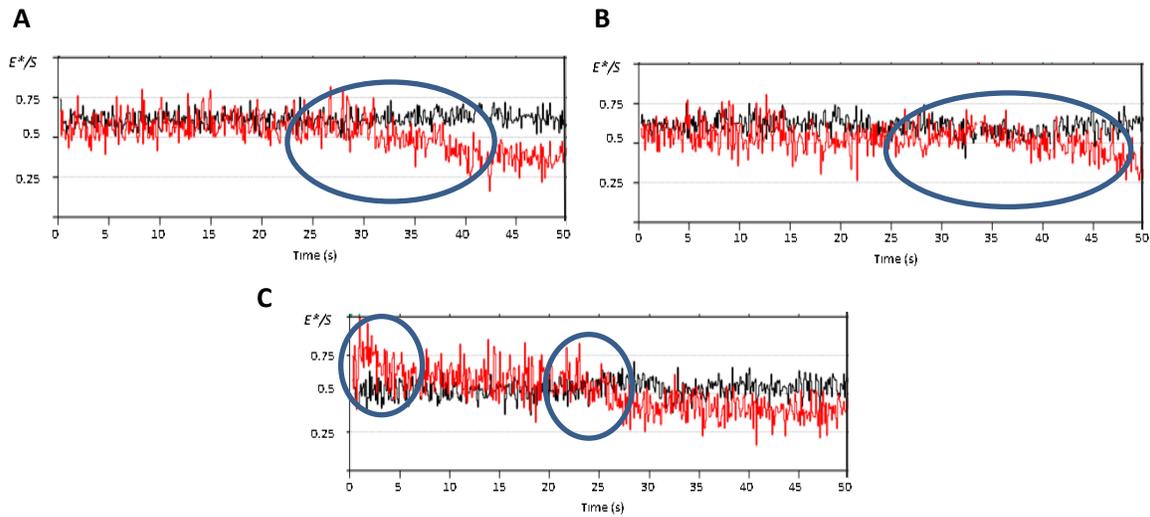
Exemplary FRET time traces representing polymerization in real time of the 004-040 DNA construct in the presence of 100 µM of each dNTP and 2 nM of HoLaMa. Polymerization (highlighted by the blue ellipses) is represented by a decrease in $E*$ to 0.28 in about 17 s (**A**), 15 s (**B**) and 2 events of 5 s each (**C**). Black line represents stoichiometry; red line represents $E*$.
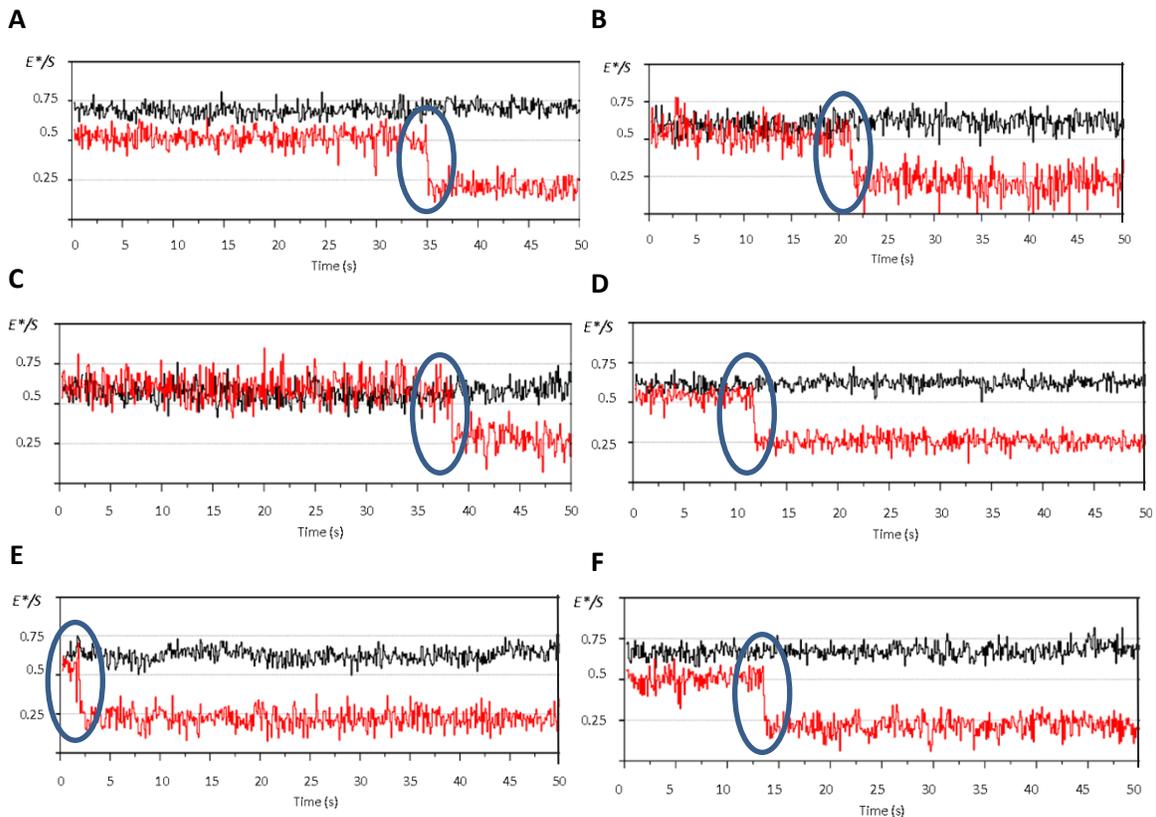
Exemplary FRET time traces representing polymerization in real time of the 004-040 DNA in the presence of 100 µM of each dNTP and 50 pM of α subunit. All observed polymerization (highlighted by the blue ellipses) are represented by a decrease in $E*$ to 0.2 in about 1 s. Black line represents S; red line represents $E*$.

TIME TRACES OF REAL TIME POLYMERIZATION
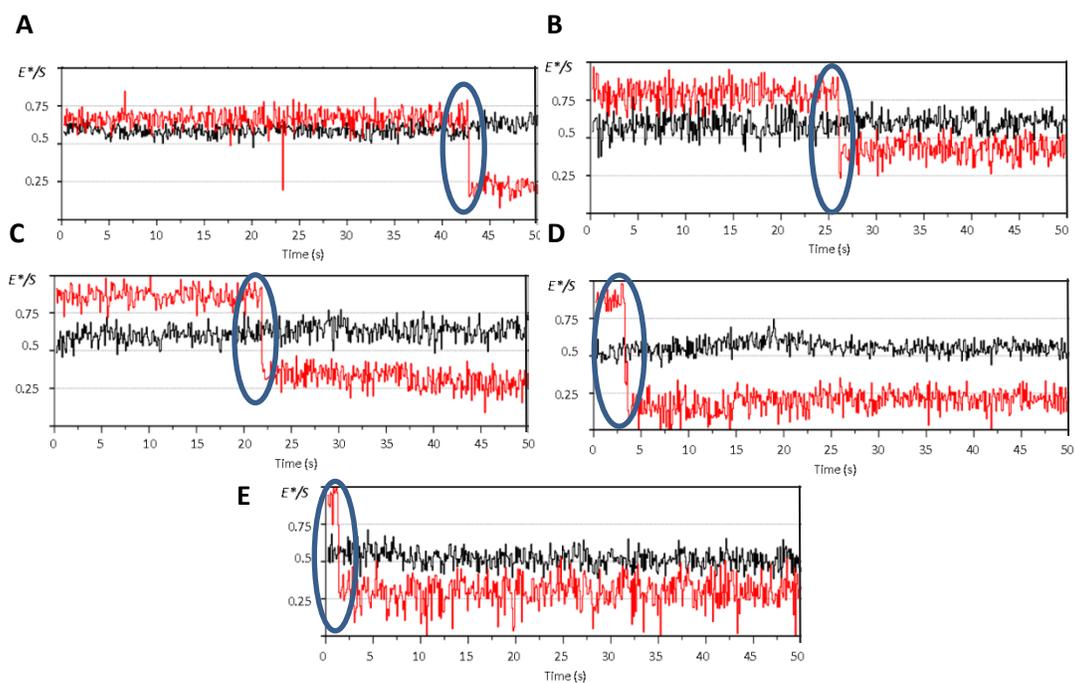
Exemplary FRET time traces representing polymerization in real time of the 004-040 DNA construct using 100 μM of each dNTP and 50 pM of αεθ. Polymerization (highlighted by the blue ellipses) is represented by a decrease in $E*$ to 0.2 in less than 1 s in all traces. Black line represents stoichiometry; red line represents $E*$.

REFERENCES

Anderson, S.G., Williams, C.R., O'Donnell, M., Bloom, L.B. (2007). A function for the ψ subunit in loading the *Escherichia coli* DNA polymerase sliding clamp. *The Journal of Biological Chemistry*, 282(10):7035-45.

Aravind, L., Koonin, E.V. (1998). Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Research*, 26(16):3746-52.

Bailey, S., Wing, R.A., Steitz, T.A. (2006). The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell*, 126(5):893-904.

Banach-Orlowska, M., Fijalkowska, I.J., Schaaper, R.M., Jonczyk, P. (2005). DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in *Escherichia coli. Mol. Microbiol*., 58(1):61-70.

Baños, B., Lazaro, J.M., Villar, L., Salas, M., de Vega, M. (2008). Characterization of a *Bacillus subtilis* 64-kDa DNA polymerase X potentially involved in DNA repair. *J. Mol. Biol.* 284:1019-1028.

Baños, B., Villar, L., Salas, M., de Vega, M. (2010). Intrinsic apurinic/apyrimidinic (AP) endonuclease activity enables *Bacillus subtilis* DNA polymerase X to recognize, incise, and further repair abasic sites. *Proc. Natl. Acad. Sci. USA,* 107(45):19219-24.

Barros, T., Guenther, J., Kelch, B., Anaya, J., Prabhakar, A., O'Donnell, M., Kuriyan, J., Lamers, M.H. (2013). A structural role for the PHP domain in *E. coli* DNA polymerase III. *BMC Struct. Biol*., 14:13-18.

Baykov, A.A., Fabrichniy, I.P., Pohjanjoki, P., Zyryanov, A.B. Reijo, L. (2000). Fluoride effects along the reaction pathway of pyrophosphatase: evidence for a second enzyme-pyrophosphate intermediate. *Biochemistry,* 39:11939-47.

Beard, W.A., Shock, D.D., Vande Berg, B.J., Wilson, S.H. (2002). Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. *J. Biol. Chem.* 277:47393-47398.

Bedford, E., Tabor, S., Richardson, C.C. (1997). The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proc Natl. Acad. Sci. USA,* 94:479-484.

Beese, L.S., Friedman, J.M., Steitz, T.A. (1993a). Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry*, 32(51):14095-101.

Beese, L. S., Derbyshire, V., Steitz, T. A. (1993b). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* 260:352-355.

Berardini, M., Foster, P.L., Loechler, E.L. (1999). DNA polymerase II (polB) is involved in a new DNA repair pathway for DNA interstrand cross-links in *Escherichia coli. J Bacteriol.* 181:2878–2882.

Berezhna, S.Y., Gill, J.P., Lamichhane, R., Millar, D.P. (2012). Single-molecule förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* 134:11261-11268.

Bessman, M.J., Kornberg, A., Lehman, I.R., Simms, E.S. (1956). Enzymic synthesis of deoxyribonucleic acid. *Biochim. Biophys. Acta.* 21(1):197-8.

Bhamre, S., Gadea, B.B., Koyama, C.A., White, S.J., Fowler, R.G. (2001). An aerobic recA-, umuC-dependent pathway of spontaneous base-pair substitution mutagenesis in *Escherichia coli. Mutat. Res.* 473:229–247.

Bierne, H., Vilette, D., Ehrlich, S. D., Michel, B. (1997). Isolation of a dnaE mutation which enhances RecA-independent homologous recombination in the *Escherichia coli* chromosome. *Molecular Microbiology*, 24(6):1225-34.

Blanco, L., Bernad, A., Lázaro, J.M., Martín, G., Garmendia, C., Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* 264: 8935-8940.

Blanco, L., Bernad, A., Salas, M. (1992). Evidence favouring the hypotesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. *Gene*, 112(1), 139-44.

Blasius, M., Shevelev, I., Jolivet, E., Sommer, S., Hubscher, U. (2006). DNA polymerase X from *Deinococcus radiodurans* possesses a structure-modulated 3'->5' exonuclease activity involved in radioresistance. *Molecular Microbiology*, 60(1):165-76.

Blinkova, A., Hervas, C., Stuckenberg, P.T., Onrust, R., O'Donell, M.E., Walker, J.R. (1993). The *Escherichia coli* DNA polymerase III holoenzyme contains both products of the dnaX gene, tau and gamma, but only tau is essential. *J. Bacteriol*. 175: 6018-6027.

Bock, C.W., Katz, A.K., Markham, G.D., Glusker, J.P. (1999). Manganese as a replacement for Magnesium and Zinc: functional comparison of the divalent ions. *J. Am. Chem. Soc.*, 121:7360-7372.

Bonner, C.A., Hays, S., McEntee, K., Goodman, M.F. (1990). DNA polymerase II is encoded by the DNA damage inducible dinA gene of *Escherichia coli. Proc. Natl. Acad. Sci. USA*, 87:7663–7667.

Boye, E., Stokke, T., Kleckner, N., & Skarstad, K. (1996). Coordinating DNA replication initiation with cell growth: Differential roles for DnaA and SeqA proteins. *Proc. Natl. Acad. Sci. USA,* 93:12206-12211.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72:248-254.

Brautigam, C.A., Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinions in Structural Biology*, 8:54–63.

Brautigam, C.A., Sun, S., Piccirilli, J.A., Steitz, T.A. (1999). Structures of normal single stranded DNA and deoxyribo-3'-S-phosphorothiolates bound to the 3'-5' exonucleolytic active site of DNA polymerase I from *Escherichia coli*. *Biochemistry*, 38(2):696-704.

Bressanin, D., Stefan, A., Piaz, F.D., Cianchetta, S., Reggiani, L., Hochkoeppler, A. (2009). Proteolysis of the proofreading subunit controls the assembly of *Escherichia coli* DNA polymerase III catalytic core. *Biochimica et Biophysica Acta*, 1794(11):1606-15.

Cai, H., Yu, H., McEntee, K., Kunkel, T.A., Goodman, M.F. (1995). Purification and properties of wild-type and exonuclease-deficient DNA polymerase II from *Escherichia coli. J. Biol. Chem*. 270:15327-15335.

Carter, J.R., Franden, M. A., Aebersold, R., Kim, D.R., McHenry, C.S. (1993). Isolation, sequencing and overexpression of the gene encoding the θ subunit of DNA polymerase III holoenzyme. *Nucleic Acids Research*, 21(14):3281-6.

Chikova, A.K., Schaaper, R.M. (2005). The Bacteriophage P1 hot Gene Product Can Substitute for the *Escherichia coli* DNA Polymerase III θ Subunit. *Journal of Bacteriology*, 187(16):5528-5536.

Christian, T.D., Romano, L.J., Rueda, D. (2009). Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc. Natl. Acad. Sci. USA*. 106:21109–21114.

Conte, E. (2012). Functions of θ subunit and the PHP domain of α subunit of *Escherichia coli* DNA Pol III. PhD Thesis, University of Bologna (Italy).

Conte, E., Vincelli, G., Shaaper, R.M., Bressanin, D., Stefan, A., Dal Piaz, F., Hochkoeppler, A. (2012). Stabilization of the *Escherichia coli* DNA polymerase III ε subunit by the θ subunit favors in vivo assembly of the Pol III catalytic core. *Arch. Biochem. Biophys*. 523(2):135-143.

Cordes, T., Vogelsang, J., Tinnefeld, P. (2009). On the mechanism of Trolox as antiblinking and antibleaching agent. *J. Am. Chem. Soc.,* 131:5018-5019.

Cowart, M., Gibson, K.J., Allen, D.J., Benkovic, S.J. (1989). DNA substrate structural requirements for the exonuclease and polymerase activities of prokaryotic and phage DNA polymerases. Biochemistry, 28:1975-1983.

Dallmann, H.G., Thimmig, R.L., McHenry, C.S. (1995). DnaX complex of *Escherichia coli* DNA polymerase III holoenzyme. Central role of tau in initiation complex assembly and in determining the functional asymmetry of holoenzyme. *J. Biol. Chem*. 270:29555-29562.

Davies, J.F., Almassy, R.J., Hostomska, Z., Ferre, R.A., Hostomsky, Z. (1994). 2.3 Å crystal structure of the catalytic domain of DNA polymerase β. *Cell* 76:1123–1133.

De Lucia, P., Cairns, J. (1969). Isolation of an *E. coli* strain with a mutation affecting DNA polymerase. *Nature,* 224(5225):1164-1166.

De Rose, E.F., Darden, T., Harvey, S., Gabel, S., Perrino, F.W., Schaaper, R.M., London, R.E. (2003). Elucidation of the ε-θ subunit interface of *Escherichia coli* DNA polymerase III by NMR spectroscopy*. Biochemistry*, 42(13):3635-44.

De Vega, M., Lázaro, J.M., Mencía, M., Blanco, L., Salas, M. (2010). Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc. Natl. Acad. Sci. USA,*107:16506-16511.

Derbyshire, V., Astatke, M., Joyce, C.M. (1993). Re-engineering the polymerase domain of Klenow Fragment and evaluation of overproduction and purification strategies. Nucleic Acids Res., 21:5439-5448.

Dohrmann, P.R., McHenry, C.S. (2005). A bipartite polymerase-processivity factor interaction: only the internal beta binding site of the alpha subunit is required for processive replication by the DNA polymerase III holoenzyme. *J. Mol. Biol.* 350(2):228-239.

Drake, J.W. (1969). Spontaneous mutation: comparative rates of spontaneous mutation. *Nature*, 221:1132-1134.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA,* 88(16):7160- 7164.

Escarceller, M., Hicks, J., Gudmundsson, G., Trump, G., Touati, D., Lovett, S., Foster, P.L., McEntee, K., Goodman, M.F. (1994). Involvement of *Escherichia coli* DNA polymerase II in response to oxidative damage and adaptive mutation. *J. Bacteriol*. 176:6221–6228.

Evans, R.J., Davies, D.R., Bullard, J.M., Christensen, J., Green, L.S., Guiles, J.W., Pata, J.D., Ribble, W. K., Janjic, N., Jarvis, T.C. (2008). Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. USA,* 105(52):20695-700.

Fabrichniy, I.P., Lehtiö, L., Salminen, A., Zyryanov, A.B., Baykov, A.A., Lahti, R., Goldman, A. (2004). Structural Studies of Metal Ions in Family II Pyrophosphatases: The Requirement for a Janus Ion. *Biochemistry*, 43:14403-14411.

Fabrichniy, I.P., Lehtiö, L., Tammenkoski, M., Zyryanov, A.B., Oksanen, E., Baykov, A.A., Lahti, R., Goldman, A. (2007). A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem.* 282:1422-1431.

Farooq, S., Fijen, C., Hohlbein, J. (2014). Studying DNA-protein interactions with single-molecule Förster Resonance Energy Transfer. *Protoplasma*, 251:317-332.

Fay, P.J., Johanson, K.O., McHenry, C.S., Barbara, R.A. (1983). Size classes of products synthesized processively by two subassemblies of *Escherichia coli* DNA polymerase III holoenzyme. *J. Biol. Chem.*, 257:5692-5699.

Fidalgo da Silva, E., Reha-Krantz, L.J. (2007). DNA polymerases proofreading: active site switching catalyzed by the bacteriophage T4 DNA polymerase. *Nucleic Acids Res.,* 35:5452.5463.

Fijalkowska, I.J., Schaaper, R.M. (1996). Mutants in the Exo I motif of *Escherichia coli* dnaQ: defective proofreading and inviability due to error catastrophe. *Proc. Natl. Acad. Sci. USA,* 93(7): 2856-61.

Fijalkowska, I.J., Dunn, R.L., Schaaper, R.M. (1997). Genetic requirements and mutational specificity of the *Escherichia coli* SOS mutator activity. *J. Bacteriol.* 179:7435–7445.

Fijalkowska, I.J., Schaaper, R.M., Jonczyk, P. (2012). DNA replication fidelity in *Escherichia coli:* a multi-DNA polymerase affair. *FEMS Microbiol .Rev.* 36(6):1105-1121.

Förster, T. (1946). Energiewanderung und Fluoreszenz. *Naturwissenschaften*, 33:166–175.

Foster, P.L., Marinus, M.G. (1992). Levels of ε, an essential replication subunit of *Escherichia coli* DNA polymerase III, are controlled by heat shock proteins. *Journal of Bacteriology*, 174(23):7509-16.

Frank, E.G., Ennis, D.G., Gonzalez, M., Levine, A.S., Woodgate, R. (1996). Regulation of SOS mutagenesis by proteolysis. *Proc. Natl. Acad. Sci. USA,* 93:10291-10296.

Freemont, P.S., Ollis, D.L., Steitz, T.A., Joyce, C.M. (1986). A domain of the Klenow Fragment of *Escherichia coli* DNA polymerase I has polymerase but not exonuclease activity. *Proteins*, 1:66-73.

Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R., Steitz, T.A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA,* 85:8924-8928.

Funabashi, H., Sigheto, H., Nakatsuka, K., Kuroda, A. (2014). A FRET-based DNA nano-tweezer technique for the imaging analysis of specific mRNA. *Analyst,* doi: 10.1039/C4AN02064B.

Gao, D., McHenry C.S. (2001a). τ Binds and Organizes *Escherichia coli* Replication Proteins through Distinct Domains. *The Journal of Biological Chemistry*, 276(6):4441-4446.

Gao, D., McHenry C.S. (2001b). τ Binds and Organizes *Escherichia coli* Replication Proteins through Distinct Domains*. The Journal of Biological Chemistry*, 276(6):4433-4440.

Georgescu, R.E., Kim, S.S., Yurieva, O., Kuriyan, J., Kong, X.P., O'Donnell, M. (2008). Structure of a sliding clamp on DNA. *Cell,* 132(1):43-54.

Georgescu, R.E., Kurth, I., O'Donnell, M. (2013). Single-molecule studies reveal the function of a third polymerase in the replisome. *Nat. Struct. Mol. Biol*., 19(1):113-116.

Georgescu, R.E., Yao, N., Indiani, C., Yurieva, O., O'Donell, M.E. (2014). Replisome mechanics: lagging strand events that influence speed and processivity. *Nucleic Acids Research*, 42(10):6497-6510.

Gleghorn, M.L., Dadydova, L.K., Basu, R., Rothman-Denes, L.B., Katsuhiko, S.M. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc. Natl. Acad. Sci USA,* 108:3566-3571.

Glover, B.P., McHenry, C.S. (2000). The DnaX-binding Subunits δ' and ψ Are Bound to γ and Not τ in the DNA Polymerase III. *The Journal of Biological Chemistry*, 275(5):3017-3020.

González, M., Frank, E.G., Levine, A.S., Woodgate, R. (1998). Lon-mediated proteolysis of the *Escherichia coli* UmuD mutagenesis protein: *in vitro* degradation and identification of residues required for proteolysis. *Genes & Development,* 12:3889–3899.

Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71:17–50.

Guex, N., Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714-2723.

Gupta, R., Hamdan, S.M., Dixon, N.E., Sheil, M.M., Beck, J.L. (2004). Application of electrospray ionization mass spectrometry to study the hydrophobic interaction between the ε and θ subunits of DNA polymerase III. *Protein Science*, 13:2878-2887.

Hamdan, S., Carr, P.D., Brown, S.E., Ollis, D.L., Dixon, N.E. (2002a). Structural basis for proofreading during replication of the *Escherichia coli* chromosome. *Structure*, 10(4):535- 46.

Hamdan, S., Bulloch, E.M., Thompson, P.R., Beck, J.L., Yang, J.Y., Crowther, J.A., Lilley, P.E., Carr, P.D., Ollis, D.L., Brown, S.E., Dixon, N.E. (2002b). Hydrolysis of the 5'-p-Nitrophenyl Ester of TMP by the Proofreading Exonuclease (ε) Subunit of *Escherichia coli* DNA Polymerase III. *Biochemistry*, 41(16):5266-5275.

Hamdan, S.M., Loparo, J.J., Takahashi, M., Richardson, C.C., van Oijen, A.M. (2009). Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* 457:336–339.

Hastings, P.J., Hersh, M.N., Thornton, P.C., Fonville, N.C., Slack, A., Frish, R.L., Ray, M.P., Harris, R.S., Leal, S.M., Rosenberg, S.M. (2010). Competition of *Escherichia coli* DNA polymerases I, II and III with DNA Pol IV in stressed cells. *PLoS One*. 5:10862.

Heikinheimo, P., Tuominen, V., Ahonen, A.K., Teplyakov, A., Cooperman, B.S., Baykov, A.A., Lahti, R., Goldman, A. (2001). Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *Proc. Natl. Acad. Sci USA,*98(6):3121-6.

Heltzel, J., Scouten Ponticelli, S.K., Sanders, L.H., Duzen, J.M., Cody, V., Pace, J., Snell, E., Sutton, M.D. (2009). Sliding Clamp-DNA interactions are required for viability and contribute to DNA Polymerase management in *Escherichia coli. Journal of Molecular Biology*, 387(1):74-91.

Hersh, M.N., Ponder, R.G., Hastings, P.J., Rosenberg, S.M. (2004). Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. *Res. Microbiol.* 155(5):352-9.

Hohlbein, J., Gryte, K., Heilemann, M., Kapanidis, A.N. (2010). Surfing on a new wave of single-molecule fluorescence methods. *Phys. Biol.* 7:031001.

Hohlbein, J., Aigrain, L., Craggs, T.D., Bermek, O., Potapova, O., Shoolizadeh, P., Grindley, N.D.F., Joyce, C.M., Kapanidis, A.N. (2013). Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun*. 4:2131-42.

Hohlbein, J., Craggs, T.D., Cordes, T. (2014). Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* 43:1156-71.

Holden, S.J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O.J., Kapanidis, A.N. (2010). Defining the limits of single-molecule FRET resolution in TIRF microscopy. *Cell* 99:3102-3111.

Horiuchi, T., Maki, H., Maruyama, M., Sekiguchi, M. (1981). Identification of the dnaQ gene product and location of the structural gene for RNase H of *Escherichia coli* by cloning of the genes. *Proc. Natl. Acad. Sci. USA.* 78:3770–3774.

Hours, C., Denhardt, D.T. (1979). Nick translation in *Escherichia coli* rep strains deficient in DNA Polymerase I activities*. Mol. Gen. Genet*. 172(1):73-80.

Huang, Y., Braithwaite, D.K., Ito, J. (1997). Evolution of dnaQ, the gene encoding the editing 3' to 5' exonuclease subunit of DNA polymerase III holoenzyme in Gram-negative bacteria. *FEBS Letters*, 400:94-98.

Hughes, A.J., Bryan, S.K., Chen, H., Moses, R.E., McHenry, C.S. (1991). *Escherichia coli* DNA polymerase II is stimulated by DNA polymerase III holoenzyme auxiliary subunits. *J. Biol. Chem.* 266:4568–4573.

Humphrey, W., Dalke, A., Schulten, K. (1996). VMD-Visual Molecular Dynamics. *J. Molec. Graphics*, 14:33-38.

Hwang, H., Kim, H., Myong, S. (2011). Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proc. Nat. Acad. Sci. USA,* 108:7414-7418.

Indiani, C., McInerney, P., Georgescu, R., Goodman, M.F., O'Donnell, M. (2005). A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. *Mol. Cell*, 19(6): 805-15.

Jarosz, D.F., Godoy, V.G., Walker, G.C. (2007). Proficient and accurate bypass of persistent DNA lesions by DinB DNA polymerases. *Cell Cycle*, 6(7):817-22.

Jay, P.J., Johanson, K.O., McHenry, C.S., Bambara, R.A. (1981). Size Classes of Products Synthesized Processively by DNA Polymerase III and DNA Polymerase III Holoenzyme of *Escherichia coli*. *The Journal of Biological Chemistry*, 256:976-983.

Jeruzalmi, D., O'Donnell, M., & Kuriyan, J. (2001a). Crystal structure of the processivity clamp loader γ (gamma) complex *of E. coli* DNA polymerase III. *Cell*, 106(4):429-41.

Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M., Kuriyan, J. (2001b). Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. Cell, 106(4):417-28.

Johnson, D.S., Bai, L., Smith, B.Y., Patel S.S., Wang, M.D. (2007). Single Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell*, 29; 129(7):1299–1309.

Jonczyk, P., Nowicka A., Fijałkowska I.J., Schaaper R.M. and Ciesla Z. (1998). In Vivo Protein Interactions within the *Escherichia coli* DNA Polymerase III Core. *Journal of Bacteriology* 180: 1563–1566.

Joyce, C.M., Steitz, T.A. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63:777-822.

Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., Weiss, S. (2004). Fluorescence-aided molecule sorting: analysis of structures and interactions by alternating laser excitation of single molecules. *Proc. Nat. Acad. Sci. USA,*101:8936-8941.

Kapanidis, A.N., Laurence, T.A., Lee, N.K., Margeat, E., Kong, X., Weiss, S. (2005). Alternating laser excitation of single molecules. *Acc. Chem. Res.* 38:523-533.

Keller, B.G., Kobitsky, A., Jäschke, A., Nienhaus, G.U., Noé, F. (2014). Complex RNA folding kinetics revealed by single-molecule FRET and Hidden Markov models. *J. Am. Chem. Soc.,* 135:4534-4543.

Kelman, Z., O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* 64:171-200.

Keniry, M.A., Park, A.Y., Owen, E.A., Hamdan, S.M., Pintacuda, G., Otting, G., Dixon, N.E. (2006). Structure of the θ subunit of *Escherichia coli* DNA polymerase III in complex with the ε subunit. *Journal of Bacteriology*, 188(12):4464-73.

Kiel, A., Kovacs, J., Mokhir, A., Krämer, R., Herten, D.P. (2007). Direct monitoring of formation and dissociation of individual metal complexes by single-molecule fluorescence spectroscopy. *Angew. Chem. Int. Ed. Engl.* 46(18):3363-3366.

Kim, D.R., McHenry, C.S. (1996a). In vivo assembly of overproduced DNA polymerase III. Overproduction, purification, and characterization of the α, α-ε, and α-ε-θ subunits. *The Journal of Biological Chemistry*, 271(34):20681-9.

Kim, S., Dallmann, H.G., McHenry, C.S., Marians K.J. (1996b). τ couples the leading- and lagging-strand polymerases at the *Escherichia coli* DNA replication fork. *J. Biol. Chem*., 271:21406–21412.

Kim, S.R., Matsui, K., Yamada, M., Gruz, P., Nohmi, T. (2001). Roles of chromosomal and episomal dinB genes encoding DNA pol IV in targeted and untargeted mutagenesis in *Escherichia coli. Mol. Genet. Genomics.* 266:207–215.

Kirby, T.W., Harvey, S., De Rose, E.F., Chalov, S., Chikova, A.K., Perrino, F.W., Schaaper, R.M., London, R.E., Lars, C. (2006). Structure of the *Escherichia coli* DNA polymerase III ε-HOT proofreading complex. *The Journal of Biological Chemistry*, 281(50):38466-71.

Klenow, H., Hennigsen, I. (1970). Selective Elimination of the Exonuclease Activity of the Deoxyribonucleic Acid Polymerase from *Escherichia coli* B by Limited Proteolysis. *Proc. Nat. Acad. Sci. USA,* 65(1):168-175.

Knippers, R. (1970). DNA polymerase II. *Nature,* 228:1050–1053.

Kong, H., Kucera, R.B., Jack, W.E. (1993). Characterization of a DNA polymerase from the hyperthermophile Archea *Thermococcus litoralis. J. Biol. Chem*., 268:1965-1975.

Kornberg, T., Gefter M.L. (1971). Purification and DNA Synthesis in Cell-Free Extracts: Properties of DNA Polymerase II. *Proc. Nat. Acad. Sci. USA,* 68(4):761-764.

Kornberg, A., Baker, T.A. (1992). DNA Replication, 2nd edn., New York: W.H. Freeman.

Kuban, W., Banach-Orlowska, M., Bialoskorska, M., Lipowska, A., Schaaper, R.M., Jonczyk, P., Fijalkowska, I.J. (2005). Mutator phenotype resulting from DNA polymerase IV overproduction in *Escherichia coli*: preferential mutagenesis on the lagging strand. *J. Bacteriol.* 187:6862–6866.

Kurz, M., Dalrymple, B., Wijffels, G., Kongsuwan, K. (2004). Interaction of the sliding clamp β-subunuit and Hda, a DnaA-related protein. *Journal of Bacteriology,* 186(11):3508-15.

Lakowicz, J.R. (2006). Principles of fluorescence spectroscopy, 3rd edn. *Springer*, Berlin.

Lamers, M.H., Georgescu, R.G., Lee, S.G., O'Donnell M., Kuriyan, J. (2006). Crystal structure of the catalytic alpha subunit of *E. coli* replicative DNA polymerase III. *Cell,* 126(5):881-892.

Lamichhane, R., Berezhna, S.Y., Gill, J.P., Van der Schans, E., Millar, D.P. (2013). Dynamics of site switching in DNA Polymerase. *J. Am. Chem. Soc*. 135:4735-4742.

Le Reste, L., Hohlbein, J., Gryte, K., Kapanidis, A.N. (2012). Characterization of dark quencher chromophores as nonfluorescent acceptors for single-molecule FRET. *Biophysical Journal*, 102:2658-2668.

Lee, N., Francklyn, C., Hamilton, E.P. (1987). Arabinose-induced binding of AraC protein to araI2 activates de araBAD operon promoter. *Proc. Natl. Acad.Sci.USA,* 84(24):8814-8.

Leulliot, N., Cladière, L., Lecointe, F., Durand, D., Hübscher, U., van Tilbeurgh H. (2009). The family X DNA polymerase from *Deinococcus radiodurans* adopts a non-standard extended conformation. *J. Biol. Chem*. 284(18):11992-9.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., Web, W.W. (2003). Zero-modewaveguides for single-molecule analysis at high concentrations. *Science* 299:682–686.

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

Lia, J., Michel, B., Allemand, J.F. (2012). Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science*, 335(6066):328-331.

Llères, D., Swift, S., Lamond, A.I. (2007). Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging mycroscopy. Curr. Protoc. Cytom. Unit 12.10:1-19.

Lobocka, M.B., Rose, D.J., Plunkett III, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B., Blattner, F.R. (2004). Genome of bacteriophage P1. *Journal of Bacteriology*, 186:7032–7068.

Long, X., Parks, J.W., Bagshaw, C.R., Stone, M.D. (2013). Mechanical unfolding of human telomere G-quadruplex DNA probed by fluorescence and magnetic tweezers spectroscopy. *Nucleic Acid Res.,* 41(4):2746-2755.

Lopez de Saro, F.J., Georgescu, R.E., Goodman, M.F., O'Donnell, M. (2003). Competitive processivity-clamp usage by DNA polymerases during DNA replication and repair. *The EMBO journal*, 22(23):6408-18.

Lopez de Saro, F.J., Marinus, M.G., Modrich, P., O'Donnell, M. (2006). The β Sliding Clamp Binds to Multiple Sites within MutL and MutS. *The Journal of Biological Chemistry*, 281(20):14340-14349

Luo, G., Wang, M., Konigsberg, W.H., Xie, X.S. (2007). Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA,* 104:12610-12615.

Maki, H., Kornberg, A. (1985). The polymerase subunit of DNA polymerase III of *Escherichia coli*. II. Purification of the α subunit, devoid of nuclease activities. *J. Biol. Chem.,* 260:12987–12992.

Maki,H., Kornberg, A. (1987). Proofreading by DNA polymerase III of *Escherichia coli* depends on cooperative interaction of the polymerase and exonuclease subunits. *Proc. Natl. Acad. Sci. USA*, 84:4389–4392.

Maki, H., Kornberg, A. (1988). DNA polymerase III holoenzyme of *Escherichia coli*. III. Distinctive processive polymerases reconstituted from purified subunits. *J. Biol. Chem*. 263(14):6561-6569.

Maki, H., Maki, S., & Kornberg, A. (1988). DNA Polymerase III Holoenzyme of *Escherichia coli*. *Biochemistry*, 263(14): 6570-6578.

Markiewicz, R.P., Vrtis, K.B., Rueda, D., Romano, L.J. (2012). Single-molecule microscopy reveals new insights into nucleotide selection by DNA Polymerase I. *Nucleic Acids Res.* 40:7975-7984.

Martina, C.E. (2014). From Klenow Fragment to HoLaMa: A mini-DNA Polymerase design. Master Thesis, University of Bologna (Italy). Academic Year 2013-2014.

McCauley, M.J., Shokri, L., Sefcikova, J., Venclovas, Č., Beuning, P.J., Williams, M.C. (2008). Distinct Double- and Single-Stranded DNA Binding of *E. coli* Replicative DNA Polymerase III α Subunit. *ACS Chemical Biology*, 3(9):577-587.

McCulloch, S.D., Kokoska, R.J., Masutani, C., Iwai, S., Hanaoka, F., Kunkel, T.A. (2004). Preferential cis-syn thymine dimer bypass by DNA polymerase η occurs with biased fidelity. *Nature* 428: 97-100.

McHenry, C.S. (1982). Purification and characterization of DNA polymerase III'. Identification of tau as a subunit of the DNA polymerase III holoenzyme. *J. Biol. Chem*, 257:1657-1663.

McHenry, C.S., Kornberg, A. (1997). DNA polymerase III holoenzyme of *Escherichia coli*. Purification and resolution into subunits. *J. Biol. Chem.*, 252(18):6478-6484.

McHenry, C.S. (2011). DNA Replicases from a Bacterial Perspective. *Annu. Rev. Biochem.* 80:403-436.

McInerney, P., Johnson, A., Katz, F., O'Donnell, M. (2007). Characterization of a triple DNA polymerase replisome. *Mol. Cell.*, 17; 27(4):527-538.

McKenzie, G.J., Magner, D.B., Lee, P.L., Rosenberg, S.M. (2003). The dinB operon and spontaneous mutation in *Escherichia coli. J. Bacteriol.* 185:3972–3977.

Merckel, M.C., Fabrichniy, I.P., Salminen, A., Kalkkinen, N., Baykov, A., Lahti, R., Goldman, A. (2001). Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism. *Structure*, 9(4):289-97.

Merkens, L.S., Bryan, S.K., Moses, R.E. (1995). Inactivation of the 5'-3' exonuclease of *Thermus aquaticus* DNA polymerase. *Biochimica et Biophysica Acta* 1264:243-248.

Mikheikin, A., Lin, H.K., Mehta, P., Jen-Jacobson, L., Trakselis, M.A. (2009). A trimeric DNA polymerase complex increasesthe native replication processivity. *Nucleic Acids Research*, 37(21):7194-7205.

Miller, H., Perrino, F.W. (1996). Kinetic mechanism of the 3'→ 5' proofreading exonuclease of DNA polymerase III. Analysis by steady state and pre-steady state methods. *Biochemistry*, 35:12919-12925.

Minnick, D. T., Astatke, M., Joyce, C. M., Kunkel, T. A. (1996). A thumb subdomain mutant of the large fragment of Escherichia coli DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *J Biol Chem* 271:24954-24961.

Moerner, W., Fromm, D. (2003). Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev. Sci. Instrum.,* 74:3597–3619.

Mott, M.L., Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria*. Nat Rev. Microbiol.*, 5:343–354.

Mueller, G.A., Kirby, T.W., De Rose, E.F., Li, D., Schaaper, R.M., London, R.E. (2005). Nuclear Magnetic Resonance Solution Structure of the *Escherichia coli* DNA Polymerase III θ Subunit. *Journal of Bacteriology*, 187(20):7081-7089.

Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M., Ha, T. (2004). Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* 86(4):2530-2537.

Muschielok, A., Andrecka, J., Jawhari, A., Brückner, F., Cramer, P., Michaelis, J. (2008). A nano-positioning system for macromolecular structural analysis. *Nat. Methods* 5:965–971.

Nakane, S., Nakagawa, N., Kuramitsu, S., Masui, R. (2009). Characterization of DNA polymerase X from *Thermus thermophilus* HB8 reveals the PolXc and PHP domains are both required for 3'- 5' exonuclease activity. *Nucleic Acid Res.* 37:2037-2052.

Naktinis V., Onrust R., Fang L., O'Donnell M. (1995). Assembly of a Chromosomal Replication Machine: Two DNA Polymerases, a Clamp Loader, and Sliding Clamps in One Holoenzyme Particle. *J. Biol. Chem.* 270:13358-13365.

Ninio, J. (1991). Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics*, 129:957-962.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T. (2001). The Y-family of DNA polymerases. *Mol. Cell.* 8:7-8.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. (1967). Mechanism of DNA replication possible discontinuity of DNA chain growth. *Japanese Journal of Medical Science & Biology*, 20(3):255-260.

Oksanen, E. (2009). Enzyme molecular choreography: Studies of soluble inorganic pyrophosphatases. PhD thesis, University of Helsinky (Finland).

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* 313:762-766.

Olsen, T.J., Choi, Y., Sims, P.C., Gul, O.T., Corso, B.L., Dong, C., Brown, W.A., Collins, P.G., Weiss, G.A. (2013). Electronic measurements of single-molecule processing by DNA polymerase I (Klenow Fragment). *J. Am. Chem. Soc.,* 135:7855-7860.

Omi, R., Goto, M., Miyahara, I., Manzoku, M., Ebihara, A., Hirotsu, K. (2007). Crystal structure of monofunctional histidinol phosphate phosphatase from *Thermus thermophiles* HB8. *Biochemistry*, 46(44):12618-27.

Onrust, R., Finkelstein, J., Naktinis, V., Turner, J., Fang, L., O'Donnell, M. (1995). Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader and sliding clamps in one holoenzyme particle. I. Organization of the clamp loader. *The Journal of Biological Chemistry*, 270:13348–13357.

Ozawa, K., Jergic, S., Park, A.Y., Dixon, N.E, Otting, G. (2008). The proofreading exonuclease subunit ε of *Escherichia coli* DNA polymerase III is tethered to the polymerase subunit α via a flexible linker. *Nucleic Acids Research*, 36(15):5074-5082.

Ozawa, K., Horan, N.P., Robinson, A., Yagi, H., Hill, F.R., Jergic, S., Xu, Z.Q., Loscha, K.V., Li, N., Tehei, M., Oakley, A.J., Otting, G., Huber, T., Dixon, N.E. (2013). Proofreading exonuclease on a tether: the complex between the *E. coli* DNA polymerase III subunits α, epsilon ε and β reveals a higly flexible arrangement of the proofreading domain. *Nucleic. Acids. Res.,* 41(10):5354-5367.

Painter, P.R. (1973). Mutator genes and selection for the mutation rate in bacteria. *Genetics*, 79:649-660.

Parfenyev, A.N., Salminen, A., Halonen, P., Hachimori, A., Baykov, A.A., Lahti, R. (2001). Quaternary structure and metal ion requirement of family II pyrophosphatases from *Bacillus subtilis*, *Streptococcus gordonii*, and *Streptococcus mutans*. *J. Biol. Chem.,* 276:24511-24518.

Park, A.Y., Jergic, S., Politis, A., Ruotolo, B.T., Hirshberg, D., Jessop, L.L., Beck, J.L., Barsky, D., O'Donnell, M., Dixon, N.E., Robinson, C.V. (2010). A single subunit directs the assembly of the *Escherichia coli* DNA sliding clamp loader. *Structure*, 18(3):285-292.

Patel, P.H., Loeb, L.A. (2000). DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci. USA,* 97(10):5095-100.

Patel, P.H., Suzuki, M., Adman, E., Shinkai, A., Loeb, L.A. (2001). Prokaryotic DNA polymerase I: evolution, structure, and base flipping mechanism for nucleotide selection. *J. Mol. Biol.* 308(5):823-37.

Pavlov, A.R., Belova, G.I., Kozyavkin, S.A., Slesarev, A.I. (2002). Helix-hairpin-helix motifs confer salt resistance and processvity on chimeric DNA polymerases. *Proc. Natl. Acad. Sci. USA,* 99:13510-13515.

Pavlov, Y.I., Maki, S., Maki, H., Kunkel, T.A. (2004). Evidence for interplay among yeast replicative polymerase alpha, delta and epsilon from studies of exonuclease and polymerase active site mutations. *BMC Biol.* 2:11.

Perrino F., Harvey S., McNeill M. (1999). Two functional domains of the e subunit of DNA polymerase III. *Biochemistry* 38:16001-16009.

Pohlhaus, J.R., Long, D.T., O'Reilly, E., Kreuzer, K.N. (2008). The ε subunit of DNA polymerase III Is involved in the nalidixic acid-induced SOS response in *Escherichia coli. Journal of Bacteriology,* 190(15):5239-47.

Pritchard, A.E., McHenry, C.S. (1999). Identification of the acidic residues in the active site of DNA polymerase III. *Journal of Molecular Biology*, 285:1067–1080.

Quiñones, A., Kaasch, J., Kaasch, M., Messer, W. (1989). Induction of dnaN and dnaQ gene expression in *Escherichia coli* by alkylation damage to DNA. *The EMBO Journal*, 8(2):587-93.

Quiñones, A., Neumann, S. (1997). The ssb-113 allele suppresses the dnaQ49 mutator and alters DNA supercoiling in *Escherichia coli. Molecular Microbiology*, 25(2):237-46.

Rangarajan, S., Woodgate, R., Goodman, M.F. (1999). A phenotype for engmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated *Escherichia coli. Proc. Natl. Acad. Sci. USA,* 94:9224–9229.

Reha-Krantz, L.J., Woodgate, S., Goodman, M.F (2014). Engineering DNA polymerases with with maximum benefit at minimum cost. *Front. Microbiol.* 5:1-17.

Reuven, N.B., Arad, G., Maor-Shoshani, A., Livneh, Z. (1999). The mutagenesis protein UmuC is a DNA polymerase activated by UmuD', RecA and SSB and specialized for translesion synthesis. *J. Biol. Chem.* 274:31763–31766.

Richetti, M., Buc, H. (1993). *E coli* DNA polymerase I as a reverse transcriptase. *The EMBO Journal* 12(2):387-396.

Robinson, A., van Oijen, A.M. (2013). Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. Nature Reviews, 11:303-315.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., Erlich, H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324(6093):163-6.

Sale, J.E., Lehmann, A.R., Woodgate, R. (2012). Y-Family DNA polymerases and their role in tolerance of cellular DNA damage. *Nature Reviews Molecular Cell Biology*, 13:141-152.

Samygina, V.R., Moiseev, V.M., Rodina, E.V., Vorobyeva, N.N., Popov, A.N., Kurilova, S.A., Nazarova, T.I., Avaeva, S.M., Bartunik, H.D. (2007). Reversible inhibition of Escherichia coli inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies. *J Mol. Biol.,* 366:1305-17.

Santoso, Y., Joyce, C.M., Potapova, O., Le Reste, L., Hohlbein, J., Torella, J.P., Grindley, N.D., Kapanidis, A.N. (2010). Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 107:715–720.

Schaaper, R.M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli. The Journal of Biological Chemistry*, 268(32):23762-5.

Scheuermann, R.H., Echols, H. (1982). A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. USA.* 81:7747–7751.

Schlacher, K., Leslie, K., Wyman, C., Woodgate, R., Cox, M.M., Goodman, M.F. (2005). DNA Polymerase V and RecA Protein, a Minimal Mutasome. *Molecular Cell,* 17:561–572.

Sevastopoulos, C.G., Glaser, D.A. (1977). Mutator action by *Escherichia coli* strains carrying dnaE mutations. *Proc. Natl. Acad. Sci. USA,* 74:3450–3497.

Shamoo, Y., Steitz, T.A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155-166.

Shapiro, A., Rivin, O., Gao, N., Hajec, L. (2005). A homogeneous, high-throughput fluorescence resonance energy transfer-based DNA polymerase assay. *Analytical biochemistry*, 347:254-261.

Sheibe, P., Bernt, E., Bergmeyer, H.U. (1974). Uric acid: UV-assay with uricase. In H.U. Bergmeyer (Ed.), Methoden der Enzymatischen Analyse, Verlag Chemie, Weinheim. pp 1999-2005.

Shevchenko, A., Tomas, H., Havlis, J., Olse, J.V., Mann, M. (2006). In gel-digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc,* 1(6):2856-60.

Slater, S.C., Lifsics, M.R., O'Donnell, M., Maurer, R. (1994). holE, the gene coding for the θ subunit of DNA polymerase III of *Escherichia coli:* characterization of a holE mutant and comparison with a dnaQ (ε-subunit) mutant. *Journal of Bacteriology*, 176(3):815-21.

Standish, A.J., Salim, A.A., Capon, R.J., Morona, R. (2013). Dual inhibition of DNA Polymerase PolC and protein tyrosine phosphatase uncovers a novel antibiotic target. *Biochemical and biophysical research communications*, 430:167-172.

Stano, N.M., Chen, J., McHenry, C.S. (2006). A coproofreading Zn2+-dependent exonuclease within a bacterial replicase. *Nature Structural & Molecular Biology*, 13(5):458-9.

Steitz, T. A., Smerdon, S. J., Jager, J., Joyce, C. M. (1994). A unified polymerase mechanism for non homologous DNA and RNA polymerases. *Science* 266:2022-2025.

Steitz, T.A. (2006). Visualizing polynucleotide polymerase machines at work. *EMBO Journal*, 25:3458–3468

Studwell-Vaughan, P.S., O'Donnell,M. (1993). DNA polymerase III accessory proteins. V. Tetha encoded by holE. *J. Biol. Chem*., 268:11785–11791.

Stukenberg, P.T., Turner, J., O'Donnell, M. (1994). An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. *Cell*, 78:877–887.

Suárez, A.S., Stefan, A., Lemma, S., Conte., E., Hochkoeppler, A. (2012). Continuous enzyme-coupled assay of phosphate- or pyrophosphate- releasing enzymes. *Biotechniques*, 53(2):99-103.

Sutton, M.D., Opperman, T., Walker., G.C. (1999). The *Escherichia coli* SOS mutagenesis proteins UmuD and UmuD' interact physically with the replicative DNA polymerase. *Proc. Natl. Acad. Sci. USA*, 96:12373–12378.

Suzuki, M., Baskin, D., Hood, L., Loeb, L.A. (1996). Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites in vivo with the crystal structure. *Proc. Natl. Acad. Sci. USA,* 93(18):9670-5.

Taft-Benz, S.A., Schaaper R.M. (1998). Mutational analysis of the 3'→5' proofreading exonuclease of *Escherichia coli* DNA polymerase III. *Nucleic Acids Res*. 26:4005-4011.

Taft-Benz S.A. and Schaaper R.M (2004). The q Subunit of *Escherichia coli* DNA Polymerase III: a Role in Stabilizing the e Proofreading Subunit. *J. Biol. Chem.* 186:2774–2780.

Tait, R., Harris, A., Smith, D.W. (1974). DNA repair in *Escherichia coli* mutants deficient in DNA I, II and/or III. *Proc. Nat. Acad. Sci. USA* 71(3):675-679.

Takagi, H., Watanabe, M., Kakuta, Y., Kamachi, R., Numata, T., Tanaka, I., Kimura, M. (2004). Crystal structure of the ribonuclease P protein Ph1877p from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. Biochemical and Biophysical Research Communications, 319(3):787-94.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., Goodman, M.F. (1999). UmuD'(2)C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl. Acad. Sci. USA,* 96(16):8919-24.

Tang, M., Pham, P., Shen, X., Taylor, J.S., O'Donnell, M., Woodgate, R., Goodman, M.F. (2000). Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature.* 404:1014–1018.

Teplova, M., Wallace, S.T., Tereshko, V., Minasov, G., Symons, A.M., Cook, P.D., Manoharan, M., Egli, M. (**1999**). Structural origins of the exonuclease resistance of a zwitterionic RNA. *Proc. Natl. Acad. Sci. USA,* 96(25):14240-5.

Teplyakov, A., Obmolova, G., Khil, P.P., Howard, A.J., Camerini-Otero, R.D., Gilliland, G.L. (2003). Crystal structure of the *Escherichia coli* YcdX protein reveals a trinuclear zinc active site. *Proteins*, 51(2):315-8.

Timinskas, K., Balvočiūtė, M., Timinskas, A., Venclovas, Č. (2014). Comprehensive analysis of DNA polymerase III a subunits and their homologs in bacterial genomes. *Nucleic Acids Res*. 42(3):1393-1413.

Tomasiewicz, H., McHenry, C.S. (1987). Sequence analysis of the *Escherichia coli* dnaE gene. Journal of Bacteriology, 169:5735-5744.

Tompkins, J.D., Nelson, J.L., Hazel, J.C., Leugers, S.L., Stumpf, J.D., Foster, P.L. (2003). Error-prone polymerase, DNA polymerase IV, is responsible for transient hypermutation during adaptive mutation in *Escherichia coli. J. Bacteriol.* 185(11):3469-72.

Toste Rêgo, A., Holding, A.N., Kent, H., Lamers, N.H. (2013). Architecture of the Pol III-clamp exonuclease complex reveals key roles of the exonuclease subunit in processiveDNA synthesis and repair. *EMBO J.,* 32(9): 1334-1343.

Tsoi, P.Y., Zhang, X., Sui, S.F., Yang, M. (2003). Effects of DNA mismatches on binding affinity and kinetics of polymerase-DNA complexes as revealed by surface plasmon resonance biosensor. *Analyst,* 128:1169-1174.

Tsuchihashi, Z., Kornberg, A. (1990). Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. USA,* 87(7):2516-2520.

Turner, J., Hingorani, M.M., Kelman, Z., O'Donnell, M. (1999). The internal workings of a DNA polymerase clamp-loading machine. *The EMBO journal*, 18(3):771-83.

Uphoff, S., Holden, S.J., Le Reste, L., Periz, J., van de Linde, S., Heilemann, M., Kapanidis, A. (2010). Monitoring multiple distances within a single molecule using switchable FRET. *Nature Methods,* 7:831-836.

Vaisman, A., McDonald, J., Noll, S., Uston, D., Loeb, G., Goodman, M.F., Woodgate, R. (2014). Investigating the mechanisms of ribonucleotide excision repair in *Escherichia coli. Mutation Research*, 761:21-33.

Wagner, J., Gruz, P., Kim, S.R., Yamada, M., Matsui, K., Fuchs, R.P., Nohmi, T. (1999). The dinB gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol. Cell.,* 4(2):281-6.

Wagner, J., Nohmi, T. (2000). *Escherichia coli* DNA polymerase IV mutator activity: genetic requirements and mutational specificity. *J. Bacteriol.* 182:4587–4595.

Wang, Y., Prose, D.E., Mei, L., Sullivan, J.C., Finney, M., Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improve performance *in vitro*. *Nucleic Acids Research* 32:1197-1207.

Wang, J., Li, T., Guo, X., Lu, Z. (2005). Exonuclease III protection assay with FRET probe for detecting DNA binding proteins. *Nucleic Acids Research*, 33:23-31.

Watson, J. D., Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737-738.

Weiss, S. (1999). Fluorescence spectroscopy of single biomolecules. *Science*, 283:1676–1683.

Welch, M.M., McHenry, C.S. (1982). Cloning and identification of the product of the dnaE gene of *Escherichia coli. J. Bacteriol*. 152:351–356.

Wieczorek, A., McHenry, C.S. (2006). The NH2-terminal php domain of the alpha subunit of the *Escherichia coli* replicase binds the epsilon proofreading subunit. *The Journal of Biological Chemistry*, 281(18):12561-7.

Wing, R.A., Bailey, S., Steitz, T.A. (2008). Insights into the replisome from the structure of a ternary complex of the DNA polymerase III α-subunit. *Journal of Molecular Biology*, 382(4):859-69.

Witte, G., Urbanke, C., Curth, U. (2003). DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery. *Nucleic. Acids. Res.,* (15):4434-4440.

Wolff, E., Kim, M., Hu, K., Yang, H., Miller, J.H. (2004). Polymerase leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* rpoB to assess the role of polymerase IV in spontaneous mutation. *J. Bacteriol*. 186:2900–2905.

Woodgate, R., Ennis, D.G. (1991). Levels of chromosomally encoded Umu proteins and requirements for in vivo UmuD cleavage. *Mol.Gen.Genet*. 229:10-16

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

Yang, S., Li, X., Ding, D., Hou, J., Jin, Z., Yu, X., Bo, T., Li, W., Li, M. (2005). A method for filling in the cohesive ends of doubled-stranded DNA using Pfu DNA polymerase. *Biotechnol. Appl. Biochem*. 42:223-226.

Yang, F., Wang, W. (2009). Structural insight into translesion synthesis by DNA Pol II. Cell, 139:1279-1289.

Yao, N., Leu, F.P., Anjelkovic, J., Turner, J., O'Donell, M. (2000). DNA structure requirements for the Escherichia coli γ complex clamp loader and DNA Polymerase III holoenzyme. *J. Biol. Chem.* 275:11440-11450.

Young, T.W., Kuhn, N.J., Wadeson, A., Ward, S., Burges, D. & Cooke, G.D. (1998). *Bacillus subtilis* ORF *yybQ* encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase? *Microbiol*. 144:2563-2571.

Yuzhakov, A., Kelman, Z., O'Donnell, M. (1999). Trading Places on DNA – A Three-Point Switch Underlies Primer Handoff from Primase to the Replicative DNA Polymerase. *Cell*, 96:153-163.

Zhao, G., Li, J., Hu, T., Wei, H., Guan, Y. (2013). Realizing directional cloning using sticky ends produced by 3' → 5' exonuclease of Klenow Fragment. *J. Biosci*. 38(5):857-866.

Zheng, Q., Juette, M.F., Jockusch, S., Wasserman, M.R., Zhou, Z., Altman, R.B., Blanchard, S.C. (2014). Ultra-stable organic fluorophores for single-molecule research. Chem. Soc. Rev., 43:1044-1056.

Zuo, Y., Deutscher, M.P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic. Acids. Res*., 29:1017–1026.