

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

Biocatalisi Applicata e Microbiologia Industriale

Ciclo XXI

Settore scientifico disciplinare di afferenza: CHIM 11

Studio della subunità  $\epsilon$  della DNA polimerasi III di  
*Escherichia coli*: stabilità e interazione  
con la subunità polimerasica

Presentata da

Dott.ssa Daniela Bressanin

Coordinatore Dottorato

Relatore

Proff. A. Hochkoeppler

Proff. A. Hochkoeppler

**Esame finale anno 2009**

*A mia nonna*

# INDICE

<b>SCOPO DELLA TESI .....</b>	<b>5</b>
<b>INTRODUZIONE .....</b>	<b>7</b>
1. LE DNA POLIMERASI .....	7
2. LE DNA POLIMERASI DI ESCHERICHIA COLI .....	10
2.1 La DNA polimerasi I .....	10
2.2 La DNA polimerasi III .....	13
2.3 Le polimerasi di Escherichia coli e le risposte alle condizioni di stress .....	21
3. LE SUBUNITÀ DEL CORE DELLA DNA POLIMERASI III .....	29
3.1 La subunità $\alpha$ .....	29
3.2 La subunità $\epsilon$ .....	32
3.3 La subunità $\theta$ .....	35
4. STABILITÀ DELLE PROTEINE .....	36
4.1 Proteasi batteriche .....	36
4.2 Regola dell'N-terminale .....	39
4.3 Riconoscimento di specifici residui nella regione C-terminale .....	41
4.4 Gli chaperones molecolari .....	41
4.5 Keio collection e ceppi deleti per proteasi e chaperones molecolari .....	43
5. APPLICAZIONI: COSTRUZIONE DI MUTATORI .....	43
5.1 Ceppi mutatori .....	43
5.2 Produzione ed impiego di ceppi mutatori .....	44
<b>MATERIALI E METODI .....</b>	<b>46</b>
1. CEPPI BATTERICI E VETTORI DI ESPRESSIONE .....	46
1.1 Terreni di coltura .....	47
1.2 Curve di crescita .....	48
2. METODICHE GENERALI .....	48
2.1 Estrazione di DNA genomico .....	48
2.2 PCR (Polymerase Chain Reaction) .....	48
2.3 Elettroforesi in gel d'agarosio .....	48
2.4 Recupero e purificazione di DNA da gel di agarosio .....	49
2.5 Clonaggio genico .....	49
2.6 Preparazione di cellule elettrocompetenti .....	50
2.7 Protocollo di trasformazione (elettroporazione) .....	50
2.9 Screening delle colonie trasformate .....	50
2.10 Midiprep di DNA plasmidico .....	51
2.11 Estrazione delle proteine totali dalle colture cellulari .....	51
2.12 Quantificazione delle proteine estratte .....	51
2.13 Gel filtrazione .....	52
2.14 Analisi SDS-PAGE .....	52
2.15 Analisi Western blot .....	52
3. REALIZZAZIONE E IMPIEGO DEL COSTRUTTO PBAD-DNAQ .....	53
3.1 Clonaggio di dnaQ .....	53
3.2 Clonaggio di forme tronche di dnaQ .....	54
3.3 Coltura, lisi, estrazione delle proteine totali e gel filtrazione .....	54
3.4 Saggio di attività esonucleasica .....	55
3.5 Analisi SDS-PAGE e Western blot .....	56
4. IMPIEGO DI CEPPI DELETI PER PROTEASI E CHAPERONES MOLECOLARI .....	56
4.1 Ceppi provenienti dalla Collezione Keio .....	56
4.2 Preparazione della stock fagica di P1 vir contenente l'allele recA1907::cat-aadA .....	56
4.3 Trasduzione dell'allele recA1907::cat-aadA mediata dal fago P1 vir .....	56
4.4 Trasformazione con il costrutto pBAD- $\epsilon$ 243 .....	57
4.5 Coltura, lisi, estrazione di proteine, gel filtrazione e analisi Western blot .....	57
5. REALIZZAZIONE E IMPIEGO DEL COSTRUTTO PBAD-DNAE .....	57

5.1 Clonaggio di <i>dnaE</i> .....	57
5.2 Clonaggio forme tronche di <i>dnaE</i> .....	58
6. REALIZZAZIONE E IMPIEGO DEL COSTRUTTO PBAD-HIS-DNAQ .....	60
6.1 Clonaggio di <i>His- dnaQ</i> .....	60
6.2 Coltura e lisi .....	60
6.3 Cromatografia d'affinità .....	60
6.4 Spettrometria di massa.....	61
7. UTILIZZAZIONE DELL'ESTRATTO PROTEICO DI CELLULE TOP10 .....	61
7.1 Coltura, lisi e gel filtrazione .....	61
7.2 Gel bidimensionale e analisi attraverso spettrometria di massa .....	62
8. UTILIZZAZIONE DEL CEPPO DI ESCHERICHIA COLI TOP10/PBAD-AR1 .....	62
8.1 Coltura, lisi, estrazione delle proteine totali.....	63
8.2 Gel filtrazione e saggio di attività esonucleasica.....	63
<b>RISULTATI E DISCUSSIONE.....</b>	<b>64</b>
1. IMPIEGO DEL COSTRUTTO PBAD-DNAQ E DELLE FORME TRONCHE DI TALE GENE.....	64
1.1 Clonaggio di <i>dnaQ</i> e di forme tronche del gene.....	64
1.2 Gel filtrazione e saggio di attività esonucleasica.....	65
1.4 Analisi SDS-PAGE e Western blotting.....	70
2. IMPIEGO DI CEPPI DELETI PER PROTEASI E CHAPERONE MOLECOLARI .....	73
2.1 Analisi Western blot .....	74
3. REALIZZAZIONE E IMPIEGO DEL COSTRUTTO PBAD-DNAE .....	76
3.1 Clonaggio di <i>dnaE</i> e di sue forme tronche ed analisi fenotipica.....	76
4. REALIZZAZIONE E IMPIEGO DEL COSTRUTTO PBAD-HIS-DNAQ.....	79
4.1 Analisi attraverso spettrometria di massa .....	81
5. UTILIZZO DEL COSTRUTTO ANTISENSO PBAD-AR1 .....	82
5.1 Gel filtrazione e saggio di attività esonucleasica.....	82
5.2 Analisi di massa.....	85
5.3 Numero di copie di oloenzima Pol III.....	86
6. ANALISI DELL'ESTRATTO PROTEICO DI CELLULE TOP10.....	86
6.1 Coltura, lisi e gel filtrazione .....	86
6.2 Gel bidimensionale e analisi attraverso spettrometria di massa .....	87
6.3 Presenza di omodimeri di <i>DnaQ</i> .....	89
7. CURVE DI CRESCITA DI BW25113 WILD TYPE E $\Delta$ DNAK734::KAN.....	89
7.1 Fattori molecolari coinvolti nella stabilità di $\mathcal{E}$ .....	90
8. ANALISI DI SEQUENZE PROTEICHE .....	91
8.1 La subunità polimerasica .....	99
8.2 La subunità esonucleasica.....	100
8.3 Interazione tra $\alpha$ ed $\mathcal{E}$ .....	101
9. L'ASSEMBLAGGIO DELLA POL III DIPENDE DALLA PROTEOLISI DELLA SUBUNITÀ $\mathcal{E}$ .....	101
10. IL MODELLO DI SCAMBIO DELLE POLIMERASI.....	103
<b>CONCLUSIONI .....</b>	<b>105</b>
<b>BIBLIOGRAFIA.....</b>	<b>106</b>

## SCOPO DELLA TESI

Le DNA polimerasi sono complessi enzimatici responsabili della replicazione del genoma degli organismi procariotici ed eucariotici. La replicazione del DNA deve essere accurata perché le informazioni genetiche siano trasmesse in maniera precisa alle successive generazioni.

La DNA polimerasi III (Pol III) è l'enzima responsabile della replicazione del genoma di *Escherichia coli*. Essa è costituita da un complesso binario asimmetrico formato da 10 subunità distinte. Il ruolo di tale polimerasi è quello di replicare simultaneamente entrambi i filamenti parentali del cromosoma batterico. L'attività replicativa minima di Pol III è garantita dal core, ovvero il complesso minimo catalitico, costituito da 3 subunità: la subunità  $\alpha$ , dotata di attività polimerasica e codificata dal gene *dnaE*, quella esonucleasica 3'→5' preposta alla correzione delle bozze (subunità  $\epsilon$ ), codificata dal gene *dnaQ*, e quella denominata  $\theta$ , codificata dal gene *holE*. La copurificazione di queste tre subunità dimostra la loro associazione nell'oloenzima Pol III. Le tre subunità si presentano in un arrangiamento lineare  $\alpha$ - $\epsilon$ - $\theta$ , in cui  $\epsilon$  è complessata sia ad  $\alpha$  che a  $\theta$ . La subunità epsilon ( $\epsilon$ ) contribuisce alla fedeltà di replicazione del genoma.

La subunità  $\epsilon$  è una proteina di 243 amminoacidi, la cui massa molecolare è pari a 27500 Da; essa è costituita da due domini, separati da una regione cerniera detta "Q linker". Il dominio N-terminale ha attività catalitica ed è importante nell'interazione con la subunità  $\theta$  del core. Il dominio C-terminale è necessario per l'interazione con la subunità  $\alpha$ ; inoltre, sia l'interazione con  $\alpha$  che quella con  $\theta$  stimolano l'attività esonucleasica di  $\epsilon$ .

Il presente lavoro è finalizzato allo studio della subunità  $\epsilon$ . In particolare, lo scopo è quello di analizzare l'associazione di tale proteina nell'oloenzima e i fattori che possono influenzare tale processo.

Nell'ambito di questa tesi, il prodotto del gene *dnaQ* è stato studiato *in vivo* sovraesprimendo la proteina intera o alcune sue varianti contenenti una parte più o meno estesa della regione C-terminale. Sono stati pertanto analizzati diversi aspetti:

- i) Per studiare la stabilità e l'attività delle varianti proteiche è stata valutata l'attività esonucleasica di  $\epsilon$  e sono state effettuate analisi western blotting con un anticorpo policlonale in grado di riconoscere  $\epsilon$ .
- ii) Per ampliare le conoscenze relative ai meccanismi coinvolti nell'assemblaggio di  $\epsilon$  nell'oloenzima è stata studiata l'interazione con la subunità polimerasica per identificare quali sono le regioni di queste proteine che interagiscono reciprocamente.
- iii) Per identificare i fattori molecolari coinvolti nella stabilità e nella degradazione di  $\epsilon$  *in vivo*, il gene *dnaQ* è stato sovraespresso in ceppi di *E. coli* deleti per alcune proteasi o per alcuni chaperones molecolari.

Le informazioni ottenute da questa lavoro possono essere utilizzate per progettare un potenziale ceppo mutatore, nel quale la sovraespressione della porzione di  $\epsilon$  che interagisce con  $\alpha$  priverebbe la cellula della subunità enzimatica con attività di correzione delle bozze. Questo porterebbe, di conseguenza, ad un aumento della frequenza di mutazioni spontanee introdotte casualmente nel genoma.

# INTRODUZIONE

## 1. Le DNA polimerasi

La sopravvivenza di una cellula dipende in larga misura dalla sua capacità di duplicare fedelmente il proprio materiale genetico, poiché, se da un punto di vista evolutivo una mutazione occasionale può portare ad un vantaggio selettivo, la maggior parte delle mutazioni sono sfavorevoli. La sintesi del DNA deve essere quindi altamente fedele perché le informazioni genetiche siano trasmesse in maniera precisa alle successive generazioni. Se non vengono corretti, gli errori nella sintesi del DNA possono avere conseguenze deleterie poiché sono essenzialmente permanenti. Nello stesso tempo, in condizioni sfavorevoli, l'aumento di variabilità genetica in una popolazione consente ad alcuni membri di ottenere un fenotipo che li rende in grado di sopravvivere e proliferare. Un transiente aumento della frequenza di mutazione potrebbe essere allora particolarmente vantaggioso: la sintesi poco fedele del DNA può consentire che si verifichi l'evoluzione della specie generando una diversità genetica in grado di permettere la sopravvivenza di virus e di microrganismi che si trovino soggetti a cambiamenti ambientali. Successivamente gli individui sopravvissuti potrebbero ripristinare una normale frequenza di mutazione durante la replicazione del loro DNA.

Gli enzimi preposti alla replicazione e al mantenimento del genoma sono le DNA polimerasi (Pol) che copiano in direzione  $5' \rightarrow 3'$  un filamento parentale di DNA stampo producendo due molecole di DNA neosintetizzate mediante un processo semiconservativo (per una review, Kornberg e Baker, 2005). In particolare è detto filamento guida (leading strand) il filamento di DNA neosintetizzato, formato per sintesi continua nella stessa direzione della forca replicativa ( $5' \rightarrow 3'$ ) e filamento tardivo (lagging strand) quello sintetizzato in corti frammenti (detti frammenti di Okazaki) nella direzione opposta ( $3' \rightarrow 5'$ ) rispetto a quella della forca di replicazione.

Un'altra caratteristica degli enzimi polimerasici è la loro incapacità di iniziare la sintesi ex novo di catene nucleotidiche; infatti per avviare l'accrescimento di un filamento, le DNA polimerasi hanno bisogno di un breve segmento preesistente di RNA o di DNA chiamato innesco (primer). Quando le basi dell'innesco sono appaiate a quelle dello stampo, la DNA polimerasi aggiunge nucleotidi al gruppo ossidrilico libero presente all'estremità  $3'$  dell'innesco.

Il processo di sintesi del DNA prevede un attacco nucleofilo del gruppo  $3'$ -ossidrilico del nucleotide presente all'estremità  $3'$  della catena nascente al gruppo fosforico in  $\alpha$  del deossinucleoside  $5'$ -trifosfato (dNTP) successivo (figura 1)

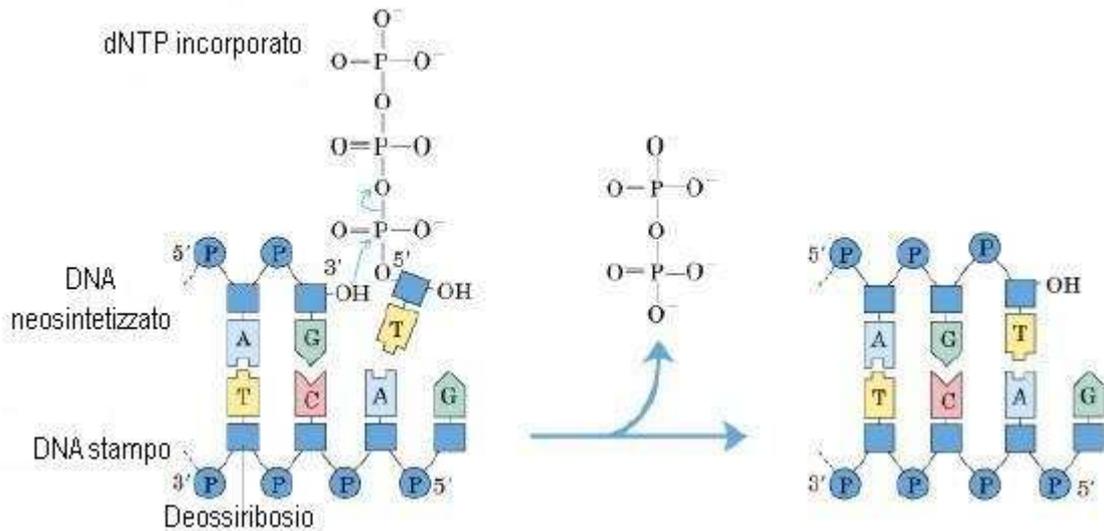
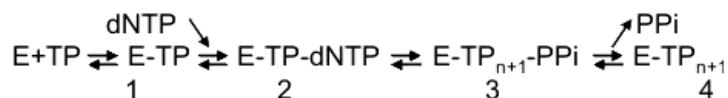


FIG 1: Allungamento di una catena di DNA. Una catena singola spaiata funziona da stampo e fornisce un terminale 3' libero al quale vengono aggiunti deossinucleoside 5'-trifosfato (dNTP). I dNTP incorporati sono selezionati sulla base della complementarità con le basi dello stampo.

Tale reazione richiede la presenza di una catena singola spaiata che funzioni da stampo ed una catena primer che fornisca un'estremità 3'-ossidrilica libera alla quale vengono aggiunte nuove unità nucleotidiche. Il filamento stampo di DNA dirige l'enzima nella selezione dello specifico nucleotide entrante che viene scelto proprio sulla base della complementarità con le basi della catena parentale, permettendo così l'aggiunta di un'unità mononucleotidica al terminale 3' della catena nascente. Il prodotto della reazione ha un nuovo gruppo ossidrilico libero in 3' e questo permette l'aggiunta del nucleotide successivo.



Nello stadio 1 l'enzima polimerasico (E) si lega allo stampo e all'innesco già sintetizzato (TP); in seguito (stadio 2) il corretto deossinucleoside 5' trifosfato (dNTP) si lega al complesso polimerasi - DNA; poi (stadio 3) l'attacco nucleofilo dá origine al legame fosfodiesterico e quindi all'incorporazione del nucleotide complementare a quello presente sullo stampo. Nella reazione si libera pirofosfato inorganico (stadio 4). L'accuratezza con la quale la DNA polimerasi replica il DNA stampo è il risultato della combinazione di tre eventi sequenziali quali la accurata incorporazione di nucleotidi, la rimozione esonucleolitica di un nucleotide aggiunto in modo non corretto e la riparazione post-replicative degli errori di appaiamento. Il nucleotide

entrante viene selezionato sulla base della specificità dell'appaiamento di basi, secondo le interazioni di Watson e Crick (per una review, Patel e Loeb, 2001). Il controllo degli errori pre-sintesi, detto "base selection" o fedeltà d'inserzione, contribuisce per un fattore  $10^4 - 10^5$  alla fedeltà di replicazione. Nei batteri si verifica una frequenza di errore molto più bassa, intorno a  $10^{-8} / 10^{-10}$  per bp replicato, perché le DNA polimerasi possono aumentare la fedeltà di replicazione mediante il controllo e la eventuale correzione del DNA neosintetizzato (proofreading) che si stima contribuiscano per un fattore  $10^2 - 10^3$  alla fedeltà di polimerizzazione. La attività proofreading consiste nella rimozione delle basi errate appena legate al filamento in corso di sintesi.

Un ultimo contributo di circa  $10^2 - 10^3$  alla fedeltà di replicazione è invece fornito dal sistema di riparo post replicativo attivato dai geni *mthLSU*, che utilizza la scarsa metilazione del filamento neo-sintetizzato per discriminare la dislocazione di basi erroneamente incorporate (sistema MMR, Methyl-directed Mismatch Repair) (per una review, Shaaper, 1993).

Dopo l'aggiunta di un nucleotide alla catena di DNA nascente, la polimerasi deve o dissociarsi o muoversi lungo lo stampo ed aggiungere un altro nucleotide. L'associazione e la dissociazione della polimerasi possono limitare la velocità complessiva della reazione, pertanto la velocità è maggiore se la polimerasi aggiunge nucleotidi in maniera continua senza dissociarsi dallo stampo. Il numero medio di nucleotidi aggiunti prima che la polimerasi si dissocia viene definito processività ed è una caratteristica unica delle DNA polimerasi. Questi enzimi hanno infatti la capacità di rimanere associati con il substrato primer-stampo nel corso di diversi cicli catalitici. La processività può influenzare drammaticamente la fedeltà di replicazione nel caso in cui, per esempio, un DNA stampo contenga un tratto omopolimerico; una bassa processività potrebbe alterare la lunghezza di tale sequenza. Le DNA polimerasi sono caratterizzate da processività molto variabili poiché alcune aggiungono soltanto pochi nucleotidi e altre ne aggiungono migliaia prima di dissociarsi. Molte DNA polimerasi necessitano di un fattore proteico aggiuntivo, non coinvolto nell'attività catalitica vera e propria, per rimanere associate allo stampo di DNA durante il corso della replicazione del genoma. Ad esempio la polimerasi del fago T7, utilizza la proteina tioredossina di *E. coli* per impedire la dissociazione dell'enzima dal DNA.

Sia nelle cellule dei procarioti che in quelle degli eucarioti sono presenti diversi tipi di DNA polimerasi. Alcune contribuiscono alla sintesi ex novo di DNA prima della divisione cellulare, altre partecipano alla riparazione e alla ricombinazione delle molecole di DNA.

Le DNA polimerasi DNA dipendenti sono state classificate in sei diverse famiglie (A, B, C, D, X, Y) sulla base dell'analisi delle loro sequenze aminoacidiche (Ohmori et al., 2001). I membri delle famiglie A, B, C, D partecipano direttamente alla replicazione del DNA mentre quelli delle famiglie X e Y sono coinvolti nei processi di riparazione del DNA. Comunque tutti i membri delle sei famiglie realizzano la polimerizzazione del DNA. Il dominio polimerasico ricorda la forma di una mano destra umana semi-aperta rivolta verso l'alto in

posizione concava ed è formata da tre domini, che, per analogia con la mano, sono detti Palma (Palm), Dita (Fingers) e Pollice (Thumb) (per una review, Patel e Loeb, 2001). Il dominio Fingers interagisce con il nucleotide entrante, il dominio Palma, che contiene il sito attivo dell'enzima, posiziona gli ioni  $Mg^{2+}$  indispensabili per la catalisi ed infine il dominio Thumb contatta il DNA a livello del solco minore. In base alla struttura secondaria del dominio Palma sono state classificate le sei famiglie delle DNA polimerasi.

Un genoma complesso, come quello umano, codifica per almeno quindici Pol, mentre quello di un organismo unicellulare come *Escherichia coli* (*E.coli*) codifica per almeno cinque diverse DNA polimerasi (per una review, O'Donnell, 2006). In particolare la Pol III replica il cromosoma batterico, le Pol II, IV e V hanno un ruolo centrale nella riparazione del DNA danneggiato e sono sotto il controllo del sistema SOS ed infine la Pol I è coinvolta sia nel processo di replicazione che in quello di riparazione.

## 2. Le DNA polimerasi di *Escherichia coli*

### 2.1 La DNA polimerasi I

Nel 1956 Arthur Kornberg e i suoi collaboratori isolarono, da un estratto proteico di cellule di *E. coli*, la DNA polimerasi I (Pol I), enzima in grado di sintetizzare DNA (Lehman et al., 1958).

La Pol I è la più abbondante polimerasi del batterio (circa 400 molecole per cellula) e, in *E. coli*, è codificata dal gene *polA* (per una review, Kornberg e Baker, 2005). L'enzima riempie le regioni vuote che si formano durante la riparazione, la ricombinazione e la replicazione del DNA. Pol I è formata da una singola catena polipeptidica costituita da 928 aminoacidi, ha una massa molecolare pari a 103.12 kDa ed è in grado di svolgere diverse funzioni. Le principali attività enzimatiche della Pol I sono la polimerizzazione e due tipologie di idrolisi esonucleolitiche indipendenti, una delle quali effettuata in direzione  $3' \rightarrow 5'$  e l'altra realizzata in direzione  $5' \rightarrow 3'$  (Setlow et al., 1972).

La polimerizzazione del DNA avviene, come per tutte le DNA polimerasi fino ad ora conosciute, in direzione  $5' \rightarrow 3'$  ed è realizzata attraverso la specificità per due substrati, ossia per l'estremità  $3'$  ossidrilica del primer e per il dNTP. Tale attività permette all'enzima di riempire i vuoti lasciati in seguito alla rimozione di vari tipi di lesioni o di basi danneggiate, come i dimeri di pirimidine indotti da radiazioni UV.

L'attività esonucleasica  $3' \rightarrow 5'$  controlla l'appaiamento delle basi a livello dell'estremità  $3'$  della catena di DNA in fase di sintesi e permette di implementare la capacità della polimerasi di incorporare il corretto dNTP durante la replicazione sulla base della specificità dell'appaiamento di basi al DNA stampo secondo le interazioni Watson e Crick. In caso di un appaiamento errato l'attività enzimatica catalizzerà l'idrolisi e la rimozione della base scorretta.

L'attività esonucleasica 5'→3' è stata riscontrata solo nella Pol I e permette la degradazione di un filamento di DNA a partire da un'estremità 5' libera (Xu et al., 1997) rilasciando così mono- o oligonucleotidi. Tale attività enzimatica permette di rimuovere per escissione sequenze di DNA neosintetizzato che contengono basi danneggiate che causano una distorsione nella struttura del DNA. Attraverso l'attività esonucleasica 5'→3' della Pol I possono quindi essere rimossi dimeri di timina formati in seguito ad esposizione a radiazioni ultraviolette (Dorson et al., 1978) e anche regioni di RNA da un doppio filamento ibrido formato da DNA e RNA. La DNA Pol I utilizza dunque la sua attività esonucleasica in 5' per rimuovere la porzione ribonucleotidica di frammenti di Okazaki e, in seguito all'escissione, l'attività DNA polimerasica 5'→3' per replicare le interruzioni generate dalla rimozione dei primers ribonucleotidici utilizzati come inneschi della replicazione del DNA. Ceppi di *E. coli* deleti per *polA* congiungono i frammenti di Okazaki con velocità 10 volte inferiore rispetto al wild type e sono sensibili ai danni al DNA causati da UV e agenti alchilanti (Okazaki et al., 1971).

La proteina codificata dal gene *polA* ha una struttura globulare con diametro di circa 65 Å ed è in grado di contattare due giri di elica di DNA (20 bp, circa 70 Å). Se tale enzima viene sottoposto a proteolisi si generano due frammenti attivi distinti: il grande frammento C-terminale di 68 kDa (detto frammento di Klenow) che possiede attività polimerasica e attività esonucleasica 3'→5' ed il piccolo frammento N-terminale di 35 kDa con attività esonucleasica 5'→3' (per una review, Kornberg e Baker, 2005). Il sito attivo della polimerasi si trova nella regione carbossi-terminale del frammento di Klenow mentre il dominio esonucleasico si trova nella regione N-terminale. La Pol I di *E. coli* è stata cristallizzata (Beese et al., 1993) e presenta la struttura tipica degli enzimi polimerasici (domini Thumb, Fingers, Palm). Il cristallo ad alta risoluzione della Pol I (Li et al., 1998) è confrontabile con altri cristalli di polimerasi appartenenti alla famiglia A che presentano una struttura tridimensionale praticamente identica. Nonostante da un punto di vista evolutivo questi enzimi siano separati da un miliardo di anni, le loro strutture risultano essere completamente sovrapponibili (per una review, Patel et al., 2001) (figura 2).

Il co-cristallo del frammento di Klenow con il DNA a doppio filamento evidenzia che il DNA è orientato in modo tale che il complesso stampo-primer si avvicini al sito catalitico dal lato adiacente il dominio esonucleasico 3'→5' (Beese et al., 1993). Studi biochimici suggeriscono che 5-8 basi del DNA a doppio filamento sono coperti dal frammento di Klenow quando l'estremità del primer è nel sito attivo polimerasico. Il dominio Fingers contatta il filamento stampo ed interagisce con il dNTP entrante (Joyce e Steitz, 1994), mentre il dominio Thumb interagisce con lo scheletro di fosfato del DNA a doppio filamento. Il frammento di Klenow catalizza la sintesi di DNA con alta fedeltà (Bebenek et al., 1990), presentando approssimativamente una frequenza pari a 10<sup>-5</sup> errori di sostituzione di una singola base per nucleotide polimerizzato e di 10<sup>-6</sup> errori di delezione di una singola base per nucleotide polimerizzato. Sebbene la subunità di proofreading contribuisca a una simile

elevata fedeltà, la maggior parte del contributo è dato dall'alta selettività della polimerasi (Minnick et al, 1996).

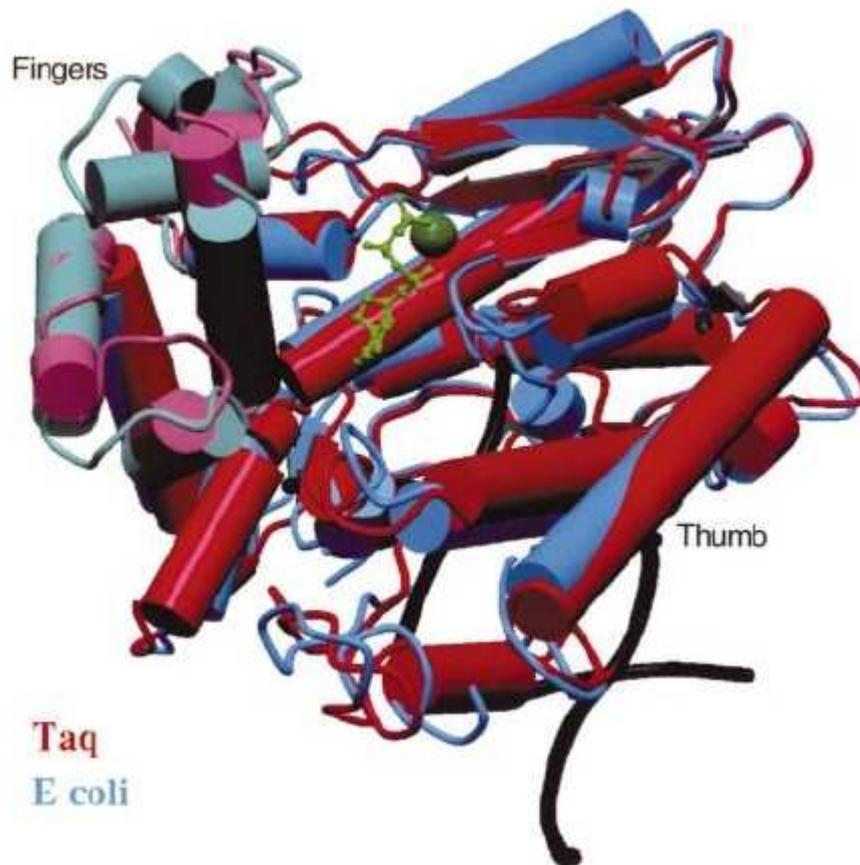


FIG 2 : Sovrapposizione delle strutture cristallizzate di Taq Pol I (in rosso) e Pol I di *E.coli* (in blu). La struttura di questi due enzimi è sovrapponibile a livello delle  $\alpha$  eliche e dei filamenti  $\beta$ . La regione corrispondente al sottodominio Finger (in magenta) differisce invece significativamente (Patel et al., 2001).

Il dominio Thumb del frammento di Klenow contiene due lunghe  $\alpha$  eliche antiparallele (dette H e I, residui 558–637) connesse da un segmento di 50 aminoacidi. Soltanto dall'analisi del co-cristallo del frammento di Klenow con il DNA (Beese et al., 1993) è stato possibile ottenere maggiori informazioni su questa regione che contiene a sua volta due corte  $\alpha$  eliche denominate  $H_1$  e  $H_2$ . Una struttura random coil connette le due piccole eliche tra di loro e con le eliche H ed I. Quando il DNA è legato si verifica un significativo cambiamento conformazionale nel dominio Thumb, così che l'estremità N-terminale dell'elica  $H_1$  sia in grado di contattare il solco minore del doppio filamento di DNA (Beese et al., 1993). Un ceppo con una delezione di 24 aminoacidi nella estremità del dominio Thumb presenta una più bassa affinità per il DNA e una ridotta processività rispetto al frammento di Klenow *wild-type*. (Minnick et al., 1996).

Pol I è anche coinvolta nel processo iniziale della replicazione dei plasmidi di tipo ColE1 che consiste nella generazione di una lunga molecola di RNA di circa 400 bp da parte della RNA polimerasi. Dopo che questo

trascritto è stato processato, viene riconosciuto dalla Pol I come primer per la sintesi di DNA a livello dell'origine di replicazione (*ori*). Successivamente, circa 400 bp a valle rispetto all' origine (segnale *n'-pas*) si verifica l'assemblaggio di un altro enzima di *E. coli*, la DNA polimerasi III (Pol III) che completa la replicazione del plasmide. In questo modo è possibile distinguere due stadi della replicazione plasmidica: una precoce, Pol I-dipendente, e una tardiva, Pol III- dipendente (Camps e Loeb, 2004).

## 2.2 La DNA polimerasi III

La DNA polimerasi III (Pol III) è l'oloenzima preposto alla replicazione del cromosoma batterico di *E.coli* (Maki et al., 1988). Tale complesso enzimatico contiene 10 subunità distinte che compongono i tre principali gruppi funzionali dell'oloenzima ovvero il core catalitico, i  $\beta$  sliding clamps e il complesso DnaX ( $\gamma_1\tau_2\delta\delta'\chi\psi$ ) che permette il caricamento dei  $\beta$  sliding clamps sul DNA (figura 3).

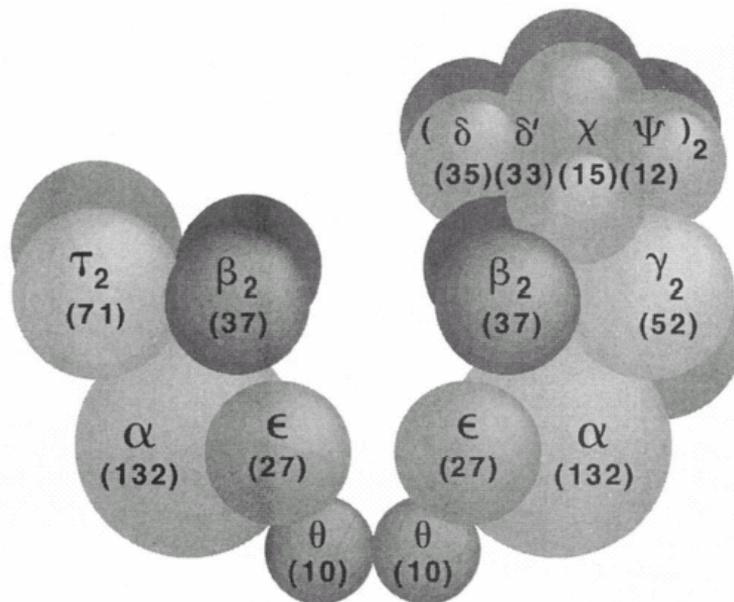


FIG 3: Struttura dell'oloenzima Pol III. Il core contiene le subunità  $\alpha$ ,  $\epsilon$  e  $\theta$  e nella Pol III sono presenti due molecole di core. Nella figura sono presenti le altre subunità che compongono l'oloenzima. Sono indicate le masse molecolari di ogni subunità (Maki et al., 1988).

L'oloenzima è un dimero asimmetrico nel quale sono presenti due distinguibili subunità polimerasiche che sintetizzano differenzialmente il filamento leading e quello lagging (Johanson e McHenry, 1984); le due molecole di core della Pol III sono tenute insieme da un dimero di  $\tau$  e sono orientate in direzioni opposte

(Glover et al., 2001). In questo modo le estremità 3' in corso di allungamento sia del filamento guida che del filamento tardivo sono vicine ma contrapposte e entrambe le polimerasi possono aggiungere desossiribonucleotidi alle due catene crescenti nello stesso momento e con la stessa velocità così che la sintesi dei due nuovi filamenti avvenga contemporaneamente.

La Pol III è un enzima altamente processivo ed è in grado di replicare efficacemente le 4.6 milioni di paia di basi del cromosoma di *E.coli* con una frequenza di errore inferiore a 1 nucleotide errato ogni  $10^{10}$  incorporati.

Un aspetto chiave della Pol III è la velocità a livello della forca replicativa che si aggira intorno a 1000 bp /s facendo 1 errore ogni  $10^5$  nucleotidi incorporati prima che avvenga la correzione proofreading (Baker e Bell, 1998).

Ogni copia del core catalitico è costituito dalle tre subunità  $\alpha$ - $\epsilon$ - $\theta$  in arrangiamento lineare nel quale  $\epsilon$  è associata sia ad  $\alpha$  che a  $\theta$ .

La subunità  $\alpha$  è il prodotto del gene *dnaE* ed è responsabile dell'attività polimerasica  $5' \rightarrow 3'$ , la subunità  $\epsilon$  è il prodotto del gene *dnaQ* il cui prodotto proteico ha attività  $3' \rightarrow 5'$  esonucleasica (proofreading) ed infine la subunità  $\theta$  è il prodotto del gene *holE* la cui funzione enzimatica non è ancora stata definitivamente chiarita. Il complesso  $\alpha$ - $\epsilon$ - $\theta$  è in grado di svolgere l'attività polimerasica minima. Nel core,  $\epsilon$  occupa la posizione centrale legando sia  $\alpha$  che  $\theta$ : la copurificazione di queste tre subunità dimostra una forte associazione fisica tra queste tre proteine nel core della Pol III.

L'aggiunta di  $\epsilon$  ad  $\alpha$  stimola l'attività polimerasica di  $\alpha$  (Maki e Kornberg, 1987), mentre l'aggiunta ad  $\epsilon$  di  $\alpha$  o di  $\theta$  ne stimola l'attività esonucleasica (Studwell-Vaughan e O'Donnell, 1993). Le interazioni fisiche e funzionali tra  $\alpha$ ,  $\epsilon$  e  $\theta$  dimostrano la reciproca influenza funzionale di queste proteine nel core della Pol III; il risultato di questa interazione è la realizzazione di una replicazione altamente fedele.

La fedeltà della replicazione è influenzata dalla subunità polimerasica  $\alpha$  e dalla subunità  $\epsilon$ , dotata di attività esonucleasica di proofreading (Taft-Benz e Schaaper, 1998).

Le subunità aggiuntive dell'oloenzima Pol III includono tra le altre  $\tau$ , che provoca la dimerizzazione dei due complessi minimi e  $\beta$ , che forma gli "sliding  $\beta$  clamps", mantiene il complesso minimo associato al DNA e ne aumenta la processività.

Tutte le polimerasi acquistano alta processività attraverso l'interazione tra la subunità catalitica e i  $\beta$  clamps che scorrono sul DNA. I  $\beta$  clamps sono caricati sul DNA dal complesso di caricamento, ATP dipendente, complesso le cui sequenze sono molto conservate tra i batteri, gli archea e gli eucarioti (per una review Bowman et al., 2005). La conservazione dei geni che codificano per le subunità del complesso di caricamento dei  $\beta$  clamps suggerisce che le proteine necessarie per la replicazione processiva hanno avuto un progenitore ancestrale comune.

La processività influenza la fedeltà della replicazione di Pol III. La subunità  $\alpha$  è accompagnata nella replicazione processiva dai  $\beta$  sliding clamps a tal punto che, in assenza di questo complesso proteico, la polimerasi replica solamente 10- 20 nucleotidi prima di staccarsi dal DNA. Anche la replicazione del DNA risulta velocizzata dalla presenza dei  $\beta$  sliding clamps, poichè il core della Pol III è in grado di incorporare nel filamento di DNA neosintetizzato soltanto 20 nucleotidi al secondo. I  $\beta$  sliding clamps sono formati da una proteina dimerica, che ha la forma di anello, avvolge lo stampo di DNA e contiene, nella regione N-terminale, due siti di legame per la subunità  $\alpha$  del core che conferiscono processività alla Pol III perchè favoriscono il contatto tra la polimerasi e il rispettivo filamento stampo (Kuriyan e O'Donnell, 1993). Tale contatto si realizza perchè i  $\beta$  sliding clamps circondano il DNA immediatamente dietro alla Pol III mediante il legame alla subunità  $\alpha$ .

Mentre la replicazione procede, i  $\beta$  sliding clamps scivolano lungo il DNA rimanendo legati al core della Pol III determinando così un aumento di velocità nell' incorporazione di nucleotidi ( $1 \text{ kb s}^{-1}$ ) e nella processività ( $> 50 \text{ kb}$ ). I  $\beta$  sliding clamps non possono assemblarsi facilmente sul DNA in maniera autonoma, ma richiedono l'intervento del complesso DnaX che, a spese di ATP, carica e blocca i  $\beta$  sliding clamps sullo stampo. Il complesso DnaX è formato da un anello pentamerico composto da proteine strutturalmente simili che appartengono alla famiglia di proteine AAA<sup>+</sup> ovvero ATPasi che partecipano ad un grande numero di attività cellulari (Jeruzalmi et al., 2001). Queste cinque proteine sono codificate dai tre geni *dnaX*, *hola* e *holB*. La subunità DnaX è l'unica delle cinque che funziona come ATPasi (Bullard et al., 2002). Il gene *dnaX* codifica per due prodotti proteici che differiscono per lunghezza:  $\gamma$  è la subunità più corta, mentre  $\tau$  è la più lunga. La subunità  $\gamma$  è generata utilizzando un modulo di lettura alternativo durante la traduzione (Flower e McHenry, 1990). Sia  $\gamma$  che  $\tau$  sono ATPasi e fungono da proteine accessorie nel caricamento dei  $\beta$  sliding clamps. In particolare, la proteina  $\tau$  contiene due domini, non presenti in  $\gamma$ , che facilitano l'interazione tra  $\tau$  e l'elicasi DnaB a livello della forca di replicazione e tra  $\tau$  e la subunità  $\alpha$  della Pol III (Gao e McHenry, 2001a,b). L'interazione tra  $\tau$  e l'elicasi DnaB aumenta più di dieci volte la velocità di svolgimento dell'elicasi da 35 a 1000 coppie di basi al secondo (Kim et al., 1996). L'estremità C-terminale della subunità  $\tau$  sporge dal complesso di caricamento dei  $\beta$  sliding clamps e, in questo modo, lega due complessi core della Pol III.  $\tau$  funziona nella forca replicativa come ponte strutturale tra l'elicasi e la polimerasi (per una review, Johnson e O'Donnell, 2005).

Il complesso DnaX include anche una copia di  $\chi$  e una copia di  $\psi$ . La subunità  $\chi$  interagisce con le proteine SSB con maggiore affinità quando SSB è legato al DNA (Glover e McHenry, 1998).  $\chi$  e  $\psi$  si associano nella replicazione del filamento lagging (Kelman et al., 1998).

### 2.2.1 La replicazione del DNA

In *Escherichia coli* la replicazione del DNA inizia a livello della regione *oriC*, l'origine di replicazione e procede in maniera bidirezionale. In questo modo si creano due forche replicative che si muovono lungo la doppia elica di DNA a partire dall'origine. Le forche replicative si muovono sul cromosoma circolare ad una velocità di circa 1000 nucleotidi al secondo e così si incontrano, dopo circa 40 minuti dall'evento iniziale, in una regione opposta ad *oriC*. In questa regione sono presenti i siti detti di terminazione o siti *Ter*, che bloccano le forche replicative (per una review, Neylon et al., 2005).

L'inizio della replicazione del DNA è un processo dinamico che vede coinvolte diverse proteine che contribuiscono all'avanzamento della forza replicativa. L'insieme di queste proteine forma il "replisoma".

Nei primi stadi dell'inizio della replicazione è coinvolto un dimero della proteina HU. Quando fu identificata, la proteina HU era stata assimilata agli istoni presenti nelle cellule eucariotiche perché mostrava caratteristiche simili quali basicità, la capacità di legare il DNA e la grande abbondanza nella cellula. Era stato pertanto ipotizzato che potesse avere un ruolo simile a quello che gli istoni hanno nelle cellule eucariotiche. In realtà, diversamente dagli istoni che favoriscono il compattamento della cromatina, HU, legandosi al DNA che presenta un superavvolgimento negativo, determina locali allentamenti transienti (Dame e Goosen, 2002) in grado quindi di favorire le tappe successive della replicazione.

In corrispondenza dell'origine di replicazione (*oriC*) avviene il riconoscimento tra copie multiple della proteina DnaA e quattro serie di sequenze, formate da nove nucleotidi ognuna, ricche in nucleotidi AT (9-mer). Questa interazione nucleoproteica e la presenza di ATP permettono la formazione e la generazione del complesso aperto (Open complex). Dopo il riconoscimento dei segmenti 9-mer, la separazione della doppia elica ha inizio a livello di altre sequenze conservate ripetute in serie e ricche in AT formate da tredici nucleotidi (13-mer) (per una review, Leonard e Grimwade, 2005). L'elicasi DnaB, costituita da un anello esamerico di subunità identiche, va ad avvolgersi attorno a ciascuna singola catena di DNA a livello della sequenza *oriC*. Il legame richiede ATP e la presenza della proteina DnaC che convoglia DnaB al complesso aperto permettendo la formazione del "complesso di preavvio" (Prepriming complex) (figura 4).

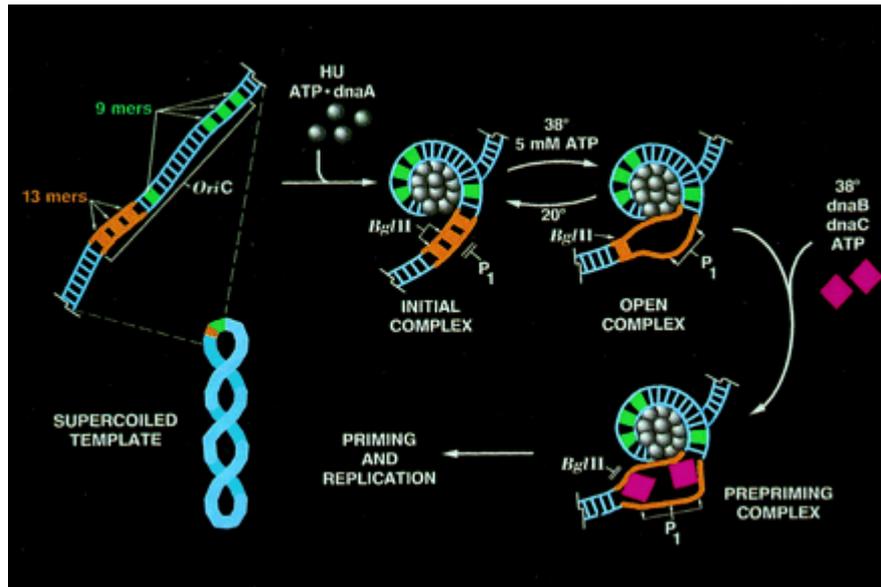


FIG 4 : Inizio della replicazione a livello dell'origine di replicazione di *Escherichia coli*. Ogni proteina DnaA lega una molecola di ATP. Un insieme di complessi formati da DnaA e ATP vanno a legare le sequenze 9-mer (initial complex). Quindi il DNA si denatura a livello delle sequenze 13-mer e si forma il complesso aperto (open complex). In seguito, l'interazione del complesso aperto con le proteine DnaB, DnaC e alla presenza di ATP permette la formazione del complesso di pre-avvio (Prepriming complex) (Kornberg, 2000).

Le regioni a singolo filamento che si formano grazie all'attività enzimatica dell'elicasi sono mantenute tali grazie al legame tra il singolo filamento di DNA e le proteine SSB che prevengono la ricostituzione della doppia elica. Poiché le DNA polimerasi cellulari non sono in grado di iniziare la sintesi di DNA in assenza di un primer, uno dei primi eventi nella sintesi di DNA è la formazione di corti inneschi di RNA. La Pol III inizia a sintetizzare DNA estendendo l'estremità 3' di un primer di RNA; la primasi DnaG (per una review, Kornberg e Baker, 2005), una RNA polimerasi, catalizza la sintesi di questi corti inneschi di RNA di circa 12 nucleotidi di lunghezza. La sintesi del filamento leading richiede solo un singolo evento enzimatico da parte della primasi, invece per la polimerizzazione discontinua del filamento lagging diversi inneschi devono essere sintetizzati ad opera di DnaG. I primers dei frammenti di Okazaki sono sintetizzati ad una distanza di circa 1000-2000 basi l'uno dall'altro; questo processo è favorito dall'interazione dinamica tra l'elicasi DnaB e la primasi DnaG (Tougu e Marians, 1996 a, b). La primasi rimane legata all'estremità 3' del primer di RNA attraverso il contatto con le proteine SSB. Il core della Pol III sostituisce poi la primasi a livello dell'estremità dell'innesco, processo a più tappe attivato dalla subunità  $\chi$  (Yuzhakov et al., 1999). Inanzitutto,  $\chi$  sostituisce la primasi legandosi in modo competitivo alle proteine SSB. Successivamente, il complesso DnaX lega l'ATP, apre l'anello dei  $\beta$  clamps e lo posiziona sul duplex RNA-DNA. Il caricamento dei  $\beta$  clamps richiede la presenza del complesso DnaX che stabilizza transientemente i  $\beta$  clamps che si trovano nella conformazione aperta permettendone così il posizionamento sul DNA. L'interazione tra i  $\beta$  clamps e il complesso DnaX può avvenire solo se

quest'ultimo lega ATP (Naktinis et al., 1995). Se si utilizza ATP o un analogo di quest'ultimo non idrolizzabile (ATP $\gamma$ S), il complesso di inizio si forma solo se  $\tau$  è presente nel complesso DnaX. Questo dato rivela che  $\alpha$  potrebbe essere assemblato nel complesso dell'oloenzima durante la reazione di trasferimento di  $\beta_2$  da parte del complesso DnaX (Dallmann et al., 1995). Inoltre, se vengono paragonate alle concentrazioni misurate in condizioni di normale presenza di ATP, in presenza di ATP $\gamma$ S, si dimezza la concentrazione di complesso iniziale che si forma. Se si aggiunge ATP $\gamma$ S a complessi di inizio purificati, formati in presenza di ATP, la metà di questi si dissocia (Glover e McHenry, 2001).

Il riconoscimento tra i  $\beta$  clamps e il complesso DnaX-ATP avviene in presenza di un DNA a singolo o doppio filamento che presenti una estremità 3' rientrante (figura 5). Solo in queste condizioni l'idrolisi di ATP è stimolata e può avvenire il caricamento dei  $\beta$  clamps sul DNA (per una review, Bowman et al., 2005). L'affinità per un'estremità 3' rientrante di DNA piuttosto che per un'estremità 5' rientrante permette alla polimerasi di iniziare immediatamente la sintesi di DNA dopo che i  $\beta$  clamps si sono complessati al DNA.

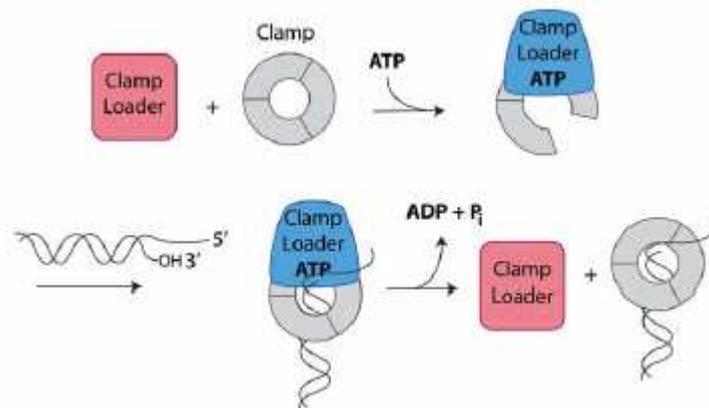


FIG 5: Il ciclo di caricamento dei  $\beta$  clamps da parte del complesso di caricamento DnaX (Clamp loader). Quando si lega l'ATP, il clamp loader acquista un'alta affinità per il filamento di DNA formato da stampo e primer che presenta un'estremità 3'libera. Il legame a tale filamento di DNA stimola l'idrolisi di ATP e il rilascio dei  $\beta$  clamps che avvolgono il DNA permettendo alla polimerasi di sintetizzare il nuovo filamento (Bowman et al, 2005).

A questo punto il core della Pol III si lega ai  $\beta$  clamps per completare il processo di sostituzione della primasi ed inizia la sintesi di un frammento di Okazaki. Quando sono legati al core della Pol III, i  $\beta$  clamps scivolano lungo il DNA con la polimerasi rendendo la Pol III un enzima processivo capace di estendere catene di DNA di più di 50 kb senza dissociarsi (Langston e O'Donnell, 2006). La forza replicativa di *E.coli* si muove di circa 1 kb al secondo in condizioni cellulari normali (Langston e O'Donnell, 2006). I frammenti di Okazaki sono lunghi 1-2 kb in *E.coli*, così i nuovi primers devono essere sintetizzati sul filamento lagging ogni 1- 2 secondi. Il

complesso DnaX è in grado di caricare, in maniera rapida e ripetuta, i nuovi dimeri di  $\beta$  clamps su i primers che sono stati appena sintetizzati dalla primasi.

L'interazione tra i  $\beta$  clamps e il core della Pol III è molto forte e i tempi di dissociazione sono maggiori di 5 minuti se la Pol III è legata al duplex primer-stampo. Dal momento che è stato calcolato che sono presenti 10-20 molecole di oloenzima nella cellula e che i nuovi frammenti di Okazaki sono prodotti in pochi secondi, le polimerasi che sintetizzano il filamento lagging dovranno essere utilizzate ripetutamente. Sono stati proposti due meccanismi per spiegare come il core della Pol III si dissocia dai  $\beta$  clamps quando finisce di sintetizzare un frammento di DNA. Secondo il "meccanismo della collisione" la subunità  $\tau$  è il segnale molecolare che attenua la forza dell'interazione tra i  $\beta$  clamps e il core della Pol III e che riduce i tempi di dissociazione da 5 minuti a 1 secondo (Leu et al., 2003). In questo modo la Pol III si libera dall'interazione con i  $\beta$  clamps e sarà pronta per sintetizzare un nuovo frammento di Okazaki. I  $\beta$  clamps, liberi dall'interazione con la Pol III possono nel frattempo legarsi a molte altre proteine come alla DNA polimerasi I (ma anche ad altre polimerasi cellulari) o alla DNA ligasi e potrebbero coordinare l'azione di queste proteine (Johnson e O'Donnell, 2005). Un'alternativa al meccanismo della collisione potrebbe essere il distacco prematuro della polimerasi dai  $\beta$  clamps prima che la sintesi del frammento di Okazaki sia completato (Li e Marians, 2000). In questo caso il segnale molecolare per il rilascio della Pol III potrebbe essere l'assemblaggio di un nuovo complesso tra i  $\beta$  clamps e un primer neosintetizzato (Yang et al., 2006).

*E. coli* è in grado di replicare il suo genoma di 4.6 Mbp a partire da una singola origine di replicazione in meno di 40 minuti e con alta accuratezza. Evidenze cellulari hanno messo in evidenza che la replicazione oltrepassa le lesioni presenti sullo stampo e lascia zone non trascritte (Wang, 2005). In passato si era ipotizzato che la sintesi dei due filamenti fosse fortemente accoppiata e il blocco di una polimerasi potesse arrestare anche l'altra finché la riparazione non fosse stata effettuata (Lee et al., 1998; Salinas e Benkovic, 2000). Studi recenti hanno invece rivelato che le condizioni che bloccano la polimerasi che sintetizza il filamento lagging non bloccano la polimerasi che sta sintetizzando il filamento leading (McInerney e O'Donnell, 2004). La sintesi del filamento leading e lagging risulta disaccoppiata anche quando le due polimerasi restano connesse ad opera del complesso DnaX attraverso le subunità  $\tau$  (Langston e O'Donnell, 2006). Quando è presente una lesione sul filamento lagging, il filamento leading continua ad essere sintetizzato e la polimerasi, bloccata sul filamento lagging, va a legarsi ad un nuovo primer, lasciando così una regione a singolo filamento con una lesione. Se la lesione è presente sul filamento leading, la polimerasi si blocca e l'elicasi recluta la primasi per riiniziare la sintesi del filamento leading al di là della lesione, lasciando una zona di DNA a singolo filamento.

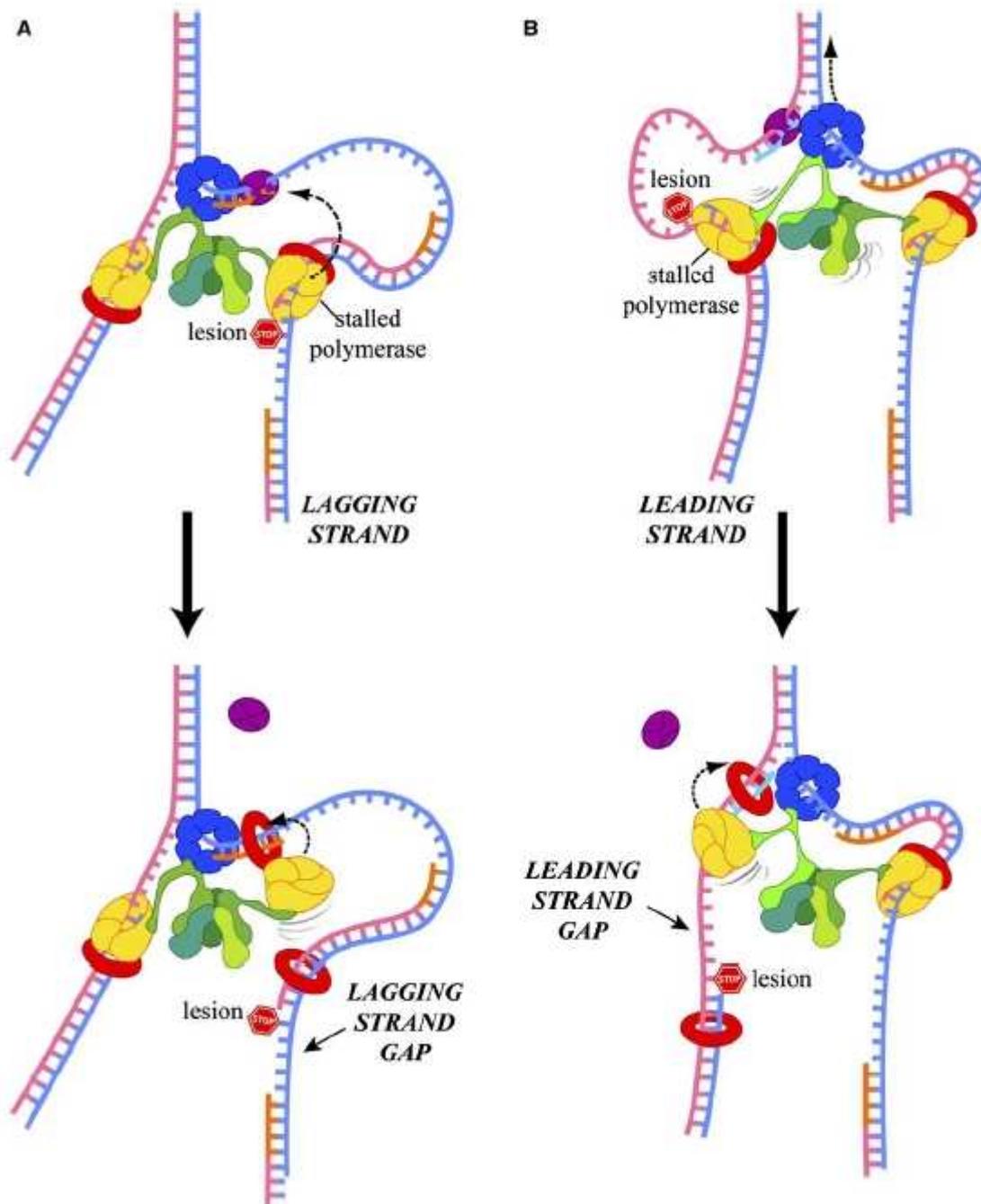


FIG 6: (A) Blocco della replicazione nel filamento lagging in caso di lesione (stop). Se la polimerasi che sta sintetizzando il filamento lagging incontra un segnale di stop, la sintesi del filamento leading continua e la polimerasi viene riciclata che si era bloccata (freccia tratteggiata) su un'altra regione dello stampo sul quale è presente il primer, lasciando una regione con una lesione a singolo filamento. (B) Blocco della replicazione nel filamento leading in caso di lesione. La polimerasi che sta sintetizzando il filamento leading si blocca quando incontra la lesione. L'elicasi recluta la primasi per riiniziare la sintesi del filamento leading superando la lesione e lasciando una regione di DNA a singolo filamento. Le regioni a singolo filamento che si sono originate potranno poi essere riparate dai sistemi di ricombinazione cellulari (da Langston e O'Donnell, 2006).

L'errore potrà essere corretto da parte dei sistemi riparativi oppure, in caso di danneggiamenti più estesi, la replicazione si arresta e si innesca una reazione di risposta allo stress.

Normalmente, se la forza replicativa non incontra ostacoli la sintesi di nuovo DNA procede fino a quando non incontra le sequenze *ter*. In *E. coli*, gli elementi *ter* arrestano l'avanzamento della forza di replicazione; essi presentano una sequenza consenso di 23 bp che costituisce il sito di legame per il prodotto del gene *tus*, una proteina di 36 kDa necessaria per la terminazione della replicazione. Quando la proteina Tus si lega alla sequenza consenso svolge un'attività controelicasica che inibisce l'azione di DnaB. Il movimento della forza viene bloccato e l'apparato replicativo si disassembla (per una review, Neylon et al., 2005).

### **2.3 Le polimerasi di *Escherichia coli* e le risposte alle condizioni di stress**

Quando il DNA di *E. coli* è soggetto ad un grave danneggiamento per esposizione a raggi UV o ad un agente mutageno si verifica un transiente aumento nella frequenza di mutazioni.

Quando i batteri sono soggetti ad un danno al DNA, circa 43 geni vengono indotti in maniera coordinata, reazione nota come "risposta SOS" (per una review, Goodman, 2002).

L'induzione dei geni SOS avviene quando il loro repressore trascrizionale comune, LexA, viene inattivato. LexA reprime i suoi geni bersaglio legandosi ad una sequenza di DNA lunga 20 bp, chiamata SOS box. Le SOS box di loci differenti sono in realtà diverse, ma esiste una sequenza consenso con 8 posizioni conservate (Courcelle et al., 2001). I geni repressi da LexA codificano per enzimi coinvolti nelle attività di riparazione del DNA. Pertanto l'induzione del sistema SOS permette alla cellula di sopravvivere nelle condizioni di avversità, per riparare o tollerare le lesioni al DNA ed eventualmente riprendere la replicazione. Durante l'induzione della risposta SOS le mutazioni vengono fissate poiché si verifica una sostituzione diretta delle basi nel DNA (per una review, Goodman, 2002).

Il termine SOS fu proposto per indicare il legame tra sopravvivenza, mutagenesi e induzione della risposta (per una review Janion, 2008). Questa cascata di eventi si verifica quando eventuali lesioni nel DNA sono così gravi da non permettere una normale replicazione da parte della Pol III.

L'inattivazione del repressore LexA è una reazione proteolitica fortemente favorita quando la proteina RecA risulta attivata (RecA\*) evento che si verifica, per esempio, dopo un solo minuto dall'irradiazione con gli UV, ma anche quando le cellule sono esposte ad altri agenti che danneggiano il DNA. In realtà, ogni volta che la concentrazione di LexA attiva diminuisce, si verifica la cascata di eventi che seguono la fine della repressione dei geni SOS. Ad esempio si è osservato in vitro che l'inattivazione di LexA aumenta quando il pH diventa leggermente alcalino (Little, 1991). In vivo LexA viene inattivata quando, in terreno ricco, le cellule

raggiungono la fase stazionaria (Dri e Moreau 1994) e in colonie invecchiate (Taddei et al, 1995). In questo modo i geni SOS potrebbero essere indotti in una grande varietà di condizioni di stress.

RecA, in seguito al danno al DNA, si lega alla regione di DNA a singolo filamento formando una struttura nucleoproteica e attivandosi. RecA\* agisce come coproteasi nel taglio autocatalitico di LexA, determinando l'induzione dei geni SOS (figura 7), (per una review, Butala et al., 2009).

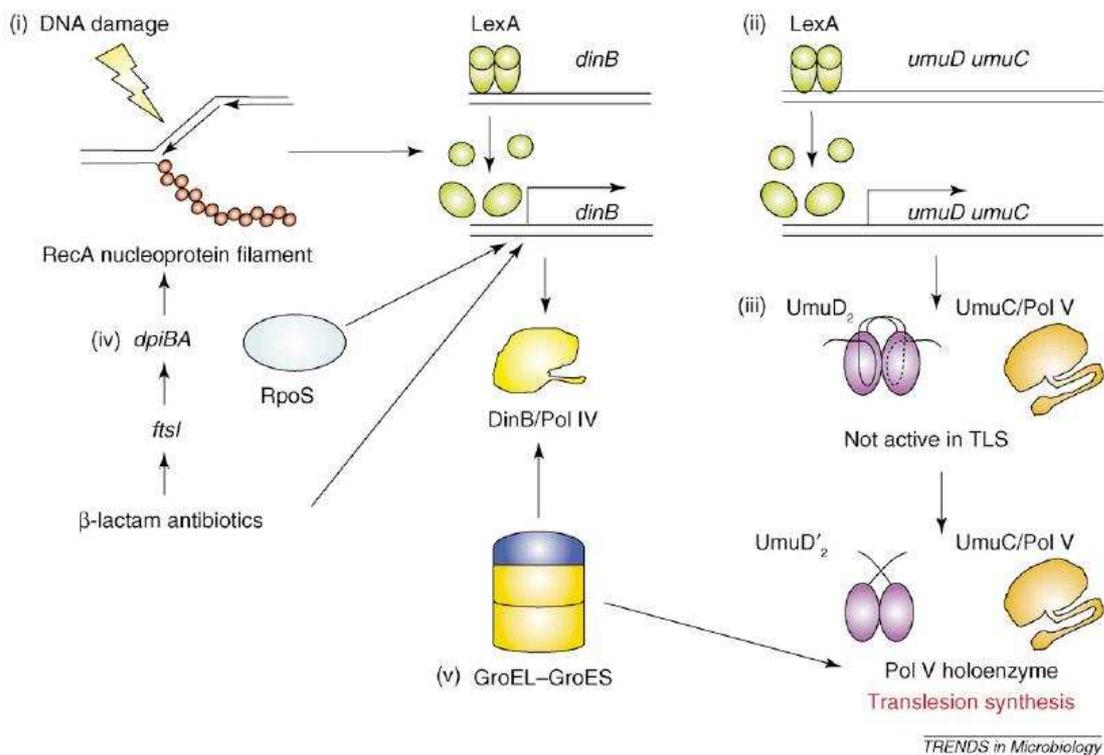


Fig 7. Induzione della risposta SOS in *E. coli*. Quando il repressore LexA è legato all' operatore a monte dei geni SOS, limita la espressione di questi geni in normali condizioni di crescita cellulare. RecA viene indotto dopo 1 minuto dall' irradiazione con gli UV formando un legame con la regione di DNA a singolo filamento. RecA\* agisce come coproteasi nel taglio autocatalitico di LexA, determinando l'induzione dei geni SOS. Nell'immagine non è presente l'induzione di Pol II (Jarosz et al., 2007).

Durante l'induzione della risposta SOS, la proteina RecA\* scinde e pertanto inattiva anche i repressori che impediscono al batteriofago  $\lambda$  di compiere il ciclo litico.

Tra i bersagli della repressione di LexA figurano i geni *dinA*, *dinB* e *umuD*'C (per una review, Courcelle et al., 2001) ovvero i geni codificanti per la DNA polimerasi II (Knippers, 1970), IV (Wagner et al; 1999) e V (Tang et al.,1999) di *E.coli*. Il livello di espressione costitutivo ed indotto di queste polimerasi è drammaticamente differente. In cellule non indotte, Pol V non è rilevabile attraverso un'analisi Western Blotting, mentre sono presenti circa 250 molecole di Pol IV per cellula. Quando viene indotta una risposta SOS sembra che siano

presenti 15 molecole di Pol V per cellula e le molecole di Pol IV siano circa 2500 (Fuchs et al., 2004). Il numero di copie di Pol II per cellula è di circa 30-50, ovvero 5 volte più elevato del numero di molecole di oloenzima Pol III (Qiu e Goodman, 1997). In seguito ad induzione della risposta SOS, il numero di copie di Pol II può raggiungere un numero 7 volte maggiore (Bonner et al., 1988).

In presenza di lesioni sul filamento leading o sul filamento lagging, la Pol III si blocca e non è in grado di superare la base danneggiata. Come precedentemente detto, la Pol III riinizierà la sintesi di DNA in un punto a valle della lesione lasciando così una zona non replicata, a singolo filamento (vedi fig 6). La formazione di una regione di DNA a singolo filamento sulla quale si va ad assemblare RecA che si attiva induce una risposta SOS quindi l'espressione delle polimerasi II, IV e V che possono accedere allo stampo di DNA danneggiato.

La Pol IV e la Pol V appartengono alla famiglia Y delle DNA polimerasi (per una review, Goodman, 2002). I membri di questa famiglia sono caratterizzati dalla capacità di replicare DNA danneggiato producendo errori (error-prone synthesis). Infatti nessuna polimerasi della famiglia Y possiede un'attività 3' → 5' esonucleasica di correzione bozze. Le Pol della famiglia Y sono caratterizzate dalla capacità di catalizzare la sintesi superando le lesioni presenti nel DNA che bloccano la replicazione. Tale processo è stato definito "translesion synthesis" (TLS); la sintesi avviene a spese di una bassa fedeltà che determina frequenti mutazioni. Questo aumento nella frequenza di mutazione determina un incremento della fitness delle cellule in crescita in un ambiente sfavorevole permettendo che si verifichi un processo di adattamento e di evoluzione. Il ruolo decisivo di queste polimerasi indica che la tolleranza ai danni nel DNA è preferibile alla morte della cellula stessa.

Un beneficio potenziale per la cellula nell'utilizzo delle polimerasi della famiglia Y è dato dalla possibilità di sostituire il complesso dell'oloenzima bloccato in caso di danno al DNA.

### 2.3.1 La DNA polimerasi II

La Pol II fu scoperta nel 1970 da Knippers (Knippers, 1970) ed è il prototipo delle polimerasi della famiglia B (Bonner et al., 1990) poiché contiene un motivo di sequenza altamente conservato in diverse DNA polimerasi di diversi regni. Si tratta di un enzima dotato, oltre che della funzione polimerasica, anche di attività esonucleasica 3' → 5' nella regione N-terminale. Come conseguenza di questa funzione, la replicazione risulta essere altamente fedele: l'enzima è in grado di sintetizzare il DNA con una frequenza di errore, per la sostituzione di una base o per una mutazione con slittamento del modulo di lettura, minore di  $10^{-6}$ . L'enzima deficiente della sua attività esonucleasica mostra invece una frequenza di errori di replicazione di 13- 240 volte maggiore rispetto alla Pol II *wild-type* (Cai et al., 1995).

La Pol II consiste di un singolo polipeptide di 89.9 kDa codificato dal gene *polB* (*dinA*) (Bonner et al., 1990). Studi genetici hanno mostrato che Pol II potrebbe essere coinvolta in una grande varietà di attività cellulari

come la riparazione del DNA danneggiato da radiazioni UV (Masker et al., 1973) permettendo una rapida (circa 10 minuti) ripresa della sintesi del DNA in seguito al trattamento (Rangarajan et al., 1999). Inoltre è stata osservata una attività di Pol II in presenza di danni ossidativi (Escarceller et al., 1994) e di legami trasversali tra filamenti (Berardini et al., 1999). Potrebbe essere inoltre coinvolta in processi di mutagenesi adattativa (per una review, Foster, 2004) e di sopravvivenza a lungo termine (Yeiser et al., 2002).

Studi in vitro hanno dimostrato che Pol II interagisce con le proteine accessorie della Pol III come i  $\beta$  clamps e il complesso per il caricamento dei  $\beta$  clamps. Interagendo con tale proteine accessorie è in grado di effettuare una sintesi di DNA più fedele e più processiva (Bonner et al., 1992).

Utilizzando il mutante *polBex1* difettivo nel locus *polB* per l'attività di proofreading, si osserva un aumento di cinque volte nella frequenza di mutazioni a livello del cromosoma di *E. coli* (Foster e Eisenstadt, 1985). Pol II, in presenza di una base incorporata in maniera errata, potrebbe partecipare alla sintesi del DNA avendo accesso al punto di crescita della replicazione. L'attività proofreading di Pol II permette dunque di mantenere l'integrità del genoma rimuovendo gli errori di replicazione (Banach-Orlowska et al., 2005). Pol II è pertanto reclutata a livello della forca di replicazione quando la Pol III è momentaneamente bloccata o non funzionale. Pol III potrebbe non essere in grado di continuare la sintesi a causa, ad esempio, di un appaiamento errato determinato da un errore nell'incorporazione di una base durante la sintesi. Del resto, studi in vitro hanno dimostrato l'incapacità della Pol III di continuare la sintesi in caso di mal appaiamento (Pham et al., 1999) e qualsiasi blocco della replicazione potrebbe rappresentare un possibile punto di entrata per altre polimerasi, come ad esempio, per la Pol II. In questo modo l'attività proofreading di Pol II può correggere gli errori effettuati dalla Pol III e rappresenta così un importante fattore di fedeltà supplementare per la replicazione di *E. coli*.

Per lo svolgimento di questa funzione finalizzata alla conservazione della sequenza nucleotidica parentale, Pol II è in grado di proteggere l'estremità 3' da sintetizzare dall'accesso della Pol IV che svolge invece la sua attività polimerasica introducendo errori ( § 2.3.2). A livello della forca replicativa, Pol II previene gli errori di replicazione (Banach-Orlowska et al., 2005). Infatti, in presenza di un nucleotide errato, la Pol III potrebbe continuare la sintesi introducendo una mutazione, correggere l'errore attraverso l'attività proofreading oppure dissociarsi. Questi tre avvenimenti non sono casuali, ma dipendono dal tipo di mutazione e dal contesto di sequenza (figura 8). In questo caso, il sito lasciato vuoto dalla Pol III, sarebbe disponibile per le altre polimerasi. Ovviamente la Pol III stessa potrebbe competere per il legame al sito. Se Pol II è presente nella cellula, processa la maggior parte dei nucleotidi mal appaiati (mismatches), ma se Pol II non è presente, la Pol IV potrà avere accesso. Pol II è quindi il maggiore processore di errori che hanno causato la dissociazione della Pol III a livello della forca replicativa e rappresenta un fattore di fedeltà nella replicazione del cromosoma di *E. coli*.

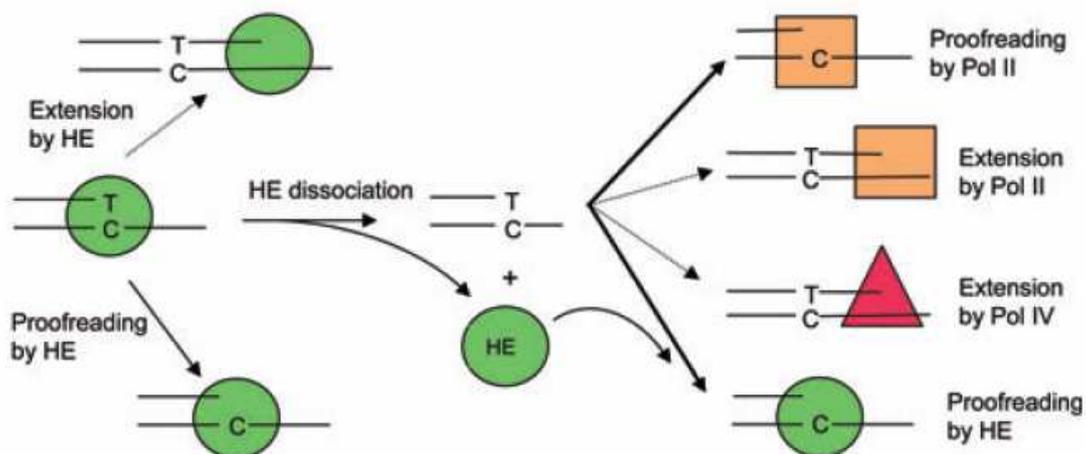


FIG 8: La figura mostra cosa potrebbe accadere qualora la DNA polimerasi III (HE, in verde) effettuasse un errore durante la sintesi di DNA (in questo caso  $G \cdot C \rightarrow T \cdot A$ ). L'appaiamento errato può essere esteso da HE causando una mutazione o rimosso per mezzo dell'attività esonucleasica di HE (proofreading). La dissociazione di HE potrebbe determinare una competizione per il sito lasciato vuoto da HE da parte delle altre polimerasi quali Pol II (quadrato arancione), Pol IV (triangolo rosso) o la stessa Pol III. Le linee più spesse indicano gli avvenimenti più probabili (Banach-Orlowska et al., 2005)

### 2.3.2 La DNA polimerasi IV

L'attività di Pol IV fu osservata per la prima volta nel 1999; l'enzima è codificato dal gene *dinB*, (damage-inducible gene) la cui espressione è indotta in caso di condizioni di stress nella cellula. I siti sul DNA per il legame a LexA nella regione a monte di *dinB* variano considerevolmente rispetto alla sequenza consenso; l'operatore a monte di *dinB* al quale si lega il repressore LexA, differisce per 7 nucleotidi rispetto alla sequenza consenso (Ohmori et al., 1995). Non essendo fortemente repressa da LexA, la Pol IV è espressa costitutivamente ad un livello basale superiore rispetto a quello delle altre polimerasi indotte in seguito a risposta SOS, ad una concentrazione di circa 250 molecole per cellula. I livelli di espressione di Pol IV aumentano poi di circa 10 volte in presenza di DNA danneggiato (Kim et al., 2001).

Si tratta di un enzima che replica poco fedelmente il DNA a causa dell'assenza di un'attività  $3' \rightarrow 5'$  esonucleasica. Nonostante l'alto livello di espressione basale, Pol IV non contribuisce significativamente alla frequenza di mutazioni nelle cellule in fase di crescita e la delezione di Pol IV causa una riduzione pari a 5-10 volte del numero di mutanti adattativi (Hersh et al, 2004). Invece è stato osservato che Pol IV è in grado di generare mutazioni frameshift negli episomi F' di popolazioni in fase stazionaria (Tompkins et al, 2003) e di essere in grado di superare il blocco replicativo della Pol III (Goodman, 2002).

Se per esempio Pol III si trovasse bloccata in seguito all'inserimento di un nucleotide errato all'estremità 3' di un primer in crescita (per una review, Kunkel e Bebenek, 2000) e tale errore non fosse corretto dalla attività esonucleasica di  $\epsilon$ , Pol IV diventerebbe indispensabile per l'estensione dell'estremità di tale primer. Questa estensione del primer contenente l'appaiamento errato genera una base non appaiata (figura 9), (Kim et al., 2001).

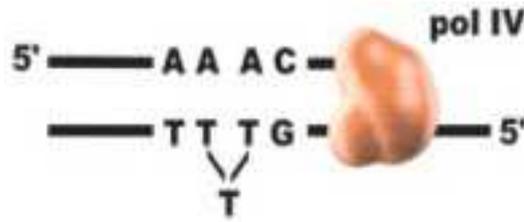


Fig 9.: L'estremità 3' di un appaiamento errato tra lo stampo e il primer viene esteso efficientemente dalla Pol IV provocando una mutazione frameshift (Goodman, 2002).

Proprio per la sua alta concentrazione cellulare è stato ipotizzato che Pol IV si leghi ai siti nei quali la replicazione è bloccata con una frequenza maggiore rispetto a quanto accade per Pol II o Pol V. In altre parole, Pol IV potrebbe essere la DNA polimerasi di *E. coli* preposta al superamento di un blocco della replicazione. Questa funzione è cruciale anche nelle cellule eucariotiche e potrebbe essere il motivo per il quale in tutti gli organismi fino ad ora conosciuti sono presenti enzimi omologhi alla Pol IV di *E. coli* (Ohmori et al., 2001). Al contrario, la Pol V è presente solo negli organismi procarioti (Ohmori et al., 2001).

L'espressione di Pol IV è indotta anche indipendentemente dall'inattivazione di LexA (Layton e Foster, 2003). Infatti, quando i batteri entrano in fase stazionaria o si trovano in condizioni di stress che determinano un blocco della crescita, un fattore sigma alternativo RpoS ( $\sigma^{38}$ ) dirige l'RNA polimerasi ai promotori di un pool di specifici geni. Il prodotto di questi geni è un insieme di proteine che permettono alla cellula di sopravvivere (per una review, Hengge-Aronis, 2002). Anche Pol IV è indotta, in tarda fase stazionaria, sotto il controllo positivo di RpoS; dopo l'induzione, sono stati registrati alti livelli della proteina per almeno tre giorni in condizioni di stress prolungato (Foster, 2004). La presenza di Pol IV è stata rilevata anche in colture di 5 giorni di età (Yeiser, 2002). Pertanto, in cellule sottoposte a stress, Pol IV può diventare la DNA polimerasi predominante, aumentando la frequenza di errori nella sintesi di DNA.

La presenza di Pol IV nella cellula è regolata anche dall'azione degli chaperones molecolari GroEL e GroES (Layton e Foster, 2005).

Qualora una popolazione microbica venga sottoposta ad una pressione selettiva non letale, si generano mutazioni. La Pol IV è coinvolta in questo processo di mutazione adattativa (Foster, 2004). Dal momento che i

batteri occupano la maggior parte della loro esistenza tentando di sopravvivere in un ambiente nel quale si sta verificando una pressione selettiva, la mutazione adattativa potrebbe giocare un importante ruolo nella sopravvivenza. Pertanto, in queste condizioni, Pol IV potrebbe agire in una sorta di equilibrio mutazionale con la Pol II (Goodman, 2002).

### 2.3.3 La DNA polimerasi V

Pol V è stata l'ultima polimerasi di *E. coli* ad essere stata identificata. L'enzima è codificato dal gene *umuD'C* (Tang et al,1999) e consiste in un dimero della proteina UmuD' modificata a livello post-traduzionale e complessata con UmuC. La Pol V è responsabile della TLS e dell'elevata frequenza di mutazione che si verifica in cellule nelle quali è stata indotta la risposta SOS. Pol V è presente nelle cellule circa 45 minuti dopo il trattamento con UV (Goodman, 2002).

Alla stregua di Pol II e di Pol IV, anche l'espressione di Pol V è regolata in risposta ad un danno al DNA.

In seguito all'induzione della risposta SOS, la proteina UmuD, che in soluzione forma omodimeri (UmuD<sub>2</sub>) e che presenta omologia di sequenza con Lexa, è processata proteoliticamente in UmuD'<sub>2</sub>. La reazione di taglio dipende dalla proteina RecA\*, che induce la latente attività autoproteolitica dell'omodimero portando alla degradazione dei 24 aminoacidi N-terminali. UmuD'<sub>2</sub> si complessa a UmuC per formare la Pol V.

Diversamente da Pol II e Pol IV, le altre due polimerasi la cui espressione è modulata in funzione di un danno al DNA, Pol V è presente nella cellula a bassa concentrazione (15 molecole per cellula) e in seguito a induzione raggiunge un numero di molecole attive pari a 200, ossia l'equivalente di molecole di Pol IV espresse a livello basale. La necessità di mantenere una bassa frequenza di mutazioni in condizioni di normale crescita cellulare significa necessariamente che le polimerasi che introducono errori devono essere mantenute strettamente controllate. La scarsità cellulare di Pol V può essere spiegata esaminando la regolazione proteolitica esercitata su UmuD. Questa proteina è degradata dalla proteasi Lon secondo un meccanismo ATP- dipendente. Il segnale per la degradazione risiede nella regione N-terminale di UmuD, regione che viene persa quando la proteina viene processata in UmuD' (Gonzalez et al., 1998). Comunque, in cellule nelle quali si è verificato un danno nel DNA, alcune molecole di UmuD riescono a sottrarsi all'azione della proteasi Lon, venendo così convertiti da RecA\* in UmuD' che è relativamente insensibile alla proteolisi Lon- dipendente e mostra in vivo un' emivita maggiore (Frank et al., 1996a).

Lon degrada anche UmuC fino a quando non viene stabilizzata da UmuD' (Donnelly e Walker, 1992).

Sia UmuD che UmuD' hanno la tendenza a formare dimeri, ma in vivo si formano in modo preferenziale eterodimeri UmuD/UmuD' (Battista et al., 1990). Tali eterodimeri sono soggetti alla proteolisi da parte di ClpXP che agisce su UmuD' e non su UmuD (Gonzalez et al., 2000). UmuD pertanto agisce come un adattatore o uno chaperone molecolare per consegnare UmuD' alle proteasi (Gonzalez et al., 2000; Neher et al., 2003a). In

ceppi deleti per la proteasi ClpXP, l'emivita dell'eterodimero UmuD/UmuD' è maggiore (Frank et al., 1996a) e la mutagenesi che insorge in seguito ad esposizione a raggi UV risulta minore probabilmente perché UmuD' viene sequestrato lontano da UmuC (Frank et al., 1996b) e non si può formare il complesso della Pol V.

Anche nella Pol V, così come per la Pol IV, manca l'attività 3' → 5' esonucleasica ed è pertanto considerata una polimerasi poco fedele (Tang et al., 2000). Nonostante la bassa concentrazione cellulare, Pol V è responsabile della maggior parte della mutagenesi indotta in seguito ad un danno al DNA. Ceppi di *E. coli* che presentano una mutazione o una delezione nel locus *UmuD* o *UmuC*, in seguito ad esposizione ad agenti mutageni o cancerogeni, presentano un fenotipo non mutatore (Woodgate, 1992). L'effetto mutatore che si presenta in seguito ad una risposta SOS riflette l'azione replicativa svolta dalla Pol V ovvero una sintesi di DNA che produce errori. Quando è presente un errore nel DNA neosintetizzato e l'oloenzima Pol III non è in grado di proseguire la sintesi di DNA e si blocca, le polimerasi accessorie come Pol V possono accedere al sito di sintesi. La DNA pol V permette il superamento del blocco, introducendo tuttavia mutazioni ad alta frequenza (figura 10); la cellula opera in questi casi una sorta di scelta: in una situazione dove il blocco di una funzione vitale quale la replicazione significa la morte, forza il sistema a procedere attraverso la produzione di diversità. Alternativamente l'effetto mutatore potrebbe essere spiegato attribuendo alla Pol V la replicazione di circa l' 1-10% del DNA del cromosoma di *E. coli* (Radman, 1999).

Il ruolo di Pol V nel conferire un fenotipo mutatore SOS-dipendente potrebbe essere così spiegata. In presenza di un errore Pol III si dissocia spontaneamente o per azione della subunità  $\tau$  (Yao e O'Donnell, 2008). A questo punto la Pol V, solo in presenza di RecA\*, può catalizzare con maggiore efficienza la sintesi del DNA, superando così le lesioni presenti nella molecola (Schlacher et al., 2005). RecA\* stimola l'attività polimerasica di Pol V di circa 300 volte (Pham et al., 2001).

Al complesso replicativo contenente la Pol V possono partecipare anche le proteine che legano il DNA a singolo filamento (SSB protein) e i  $\beta$  sliding clamps che aumentano la processività della Pol V (Schlacher et al., 2005).



FIG 10: La translesion synthesis (TLS) effettuata dalla Pol V porta alla incorporazione della base G sul filamento complementare a quello nel quale è presente un dimero di timina, causando la mutazione A → G (Goodman, 2002).

### 3. Le subunità del core della DNA polimerasi III

#### 3.1 La subunità $\alpha$

La proteina  $\alpha$  è formata da 1160 aminoacidi, costituisce la subunità catalitica della Pol III ed è codificata dal gene *dnaE*. L'enzima non mostra omologia di sequenza rispetto ad altre polimerasi. La struttura, determinata con una risoluzione di 2.3 Å, degli aminoacidi 1- 917 della proteina, mostra delle differenze rispetto alle altre famiglie di DNA polimerasi suggerendo così un'evoluzione indipendente di questa proteina.

La subunità  $\alpha$  della Pol III comprende tre domini indipendenti: il palmo (Palm domain), le dita (Finger domain) ed il pollice (Thumb domain). Questi tre domini, insieme, formano un solco profondo in grado di complessare il DNA a doppio filamento che, in questo modo, interagisce con il sito attivo della polimerasi.

Il cuore della proteina è formata dal Palm (residui 271- 432 e 511- 560), mentre il dominio Finger è localizzato nella parte sinistra rispetto al Palm ed occupa i residui da 561 a 911. Il dominio Finger è suddivisibile in quattro sottodomini che sono detti dito indice (Index finger, residui 641 - 756), dito medio (Middle finger, residui 561 – 640 e 757 – 778), dito anulare (Ring finger, residui 779 – 838 ) e dito mignolo (Little finger, residui 839 – 911). Il dominio Thumb (residui 433 – 510) è localizzato nella parte destra rispetto al dominio Palm e si inarca al di sotto di questo. Fa parte della proteina anche un ulteriore dominio detto PHP (Polymerase and Histidinol Phosphatase Domain, residui 1 - 255) inizialmente identificato grazie alla sua omologia di sequenza con l'enzima istidinol fosfatasi (figura 11).

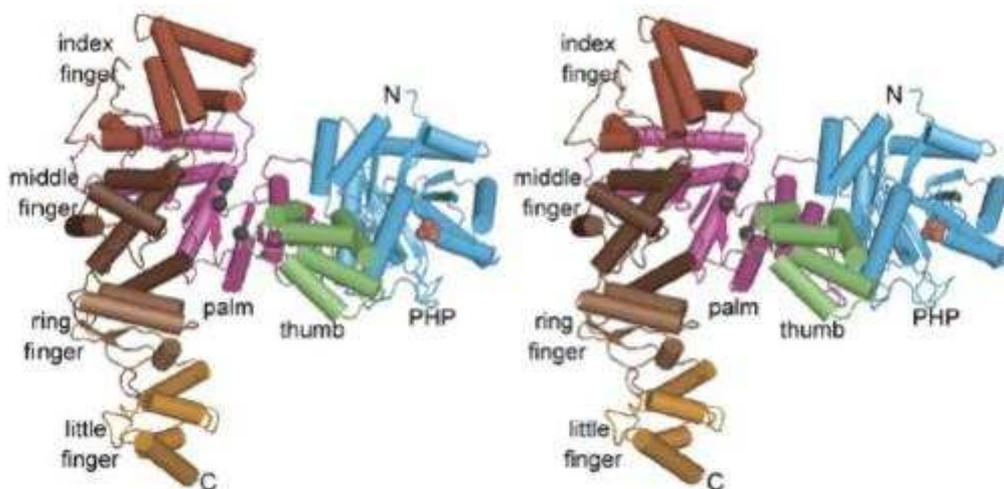


Figura 11: Struttura di  $\alpha$  (residui 1-917). Le sfere nere indicano la posizione dei residui corrispondenti al sito attivo, le sfere rosse indicano invece gli ioni fosfato nel dominio PHP (Lamers et al, 2006).

Era stato ipotizzato che il dominio PHP avesse un'attività fosfatase e, ad esempio, agisse come una pirofosfatasi in grado di idrolizzare il pirofosfato prodotto durante la sintesi del DNA (Aravind e Koonin, 1998). Recentemente è stato dimostrato che il dominio PHP di *Thermus thermophilus* possiede un'intrinseca attività 3'→5' esonucleasica e potrebbe dunque funzionare come una subunità proofreading, Zn<sup>2+</sup> dipendente, addizionale alla subunità ε, la subunità esonucleasica Mg<sup>2+</sup> dipendente di Pol III (Stano et al, 2006). PHP potrebbe essere il dominio esonucleasico ancestrale che faceva parte della proteina polimerasica e che, nel corso dell'evoluzione, è stato sostituito dall'enzima odierno codificato da un gene indipendente. Comunque PHP e ε non sono ridondanti, ma probabilmente complementari; ad esempio il dominio PHP potrebbe essere più attivo rispetto ad ε sui nucleotidi mal appaiati. In *E. coli* la delezione della regione N-terminale di α distrugge il legame alla subunità ε ed elimina la attività polimerasica ( Wieczorek e McHenry 2006 ).

L'interazione tra α e il DNA potrebbe avvenire tra i domini Thumb e Finger (figura 12 A). In particolare, il DNA passerebbe sotto il dominio Thumb che contatterebbe la molecola a livello del solco minore con diversi residui conservati e carichi positivamente (figura 12 B). Inoltre una serie di residui conservati, carichi positivamente, corre dal dominio Thumb nel dominio Palm, seguendo l'andamento della struttura del DNA. Le regioni putative di α che interagiscono con il DNA sono molto conservate ed è stato possibile identificare un sito che interagirebbe con il nucleotide entrante. Il nucleotide e lo stampo si troverebbero vicini in un loop (residui 753-758) che connette il sottodominio Index Finger al Middle Finger.

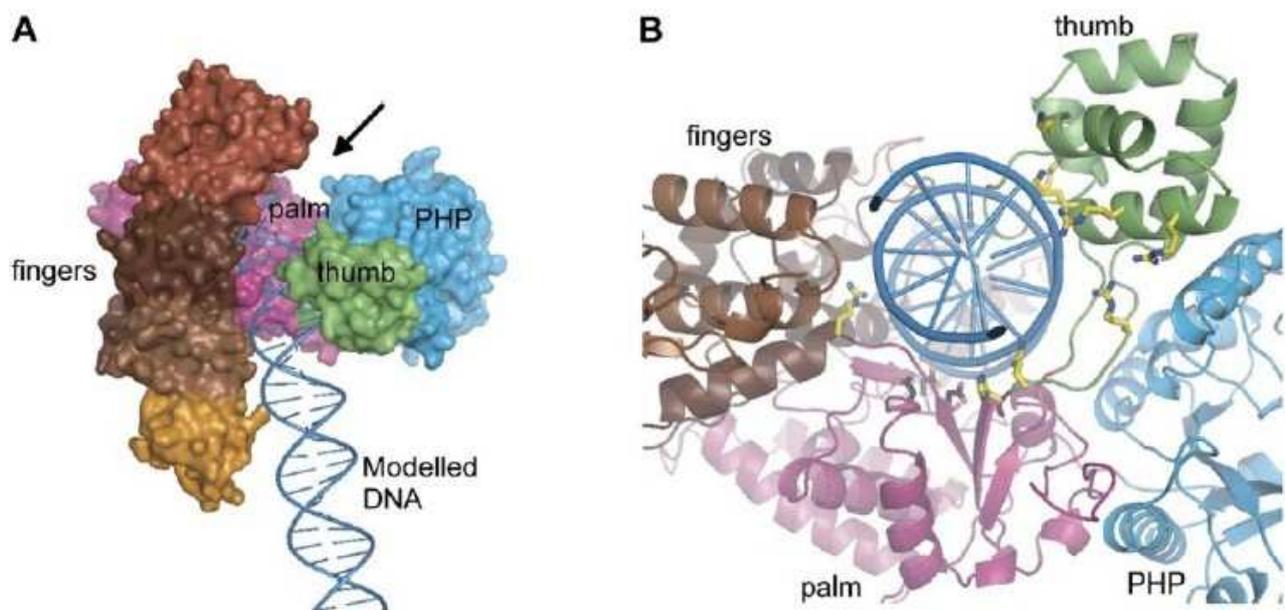


FIG 12. In A: Modello del legame tra il DNA e la subunità α che lega il DNA. B: Particolare dell'interazione tra α e il DNA. I residui in giallo sono carichi positivamente e contattano il DNA (Lamers et al, 2006).

Una serie di residui carichi positivamente sono presenti nei domini Palm e Index Finger e sono adiacenti ai tre residui di aspartato catalitici (D401, D403, D555) e favorirebbero l'interazione con l'estremità trifosfato, carica negativamente, del nucleotide entrante.

Nei residui 994-1073 è presente un dominio OB (oligonucleotide binding). Tale dominio, identificato in altre proteine, è normalmente coinvolto nell'interazione tra proteine e acidi nucleici o tra proteine e proteine (Theobald et al., 2003). Tutte le interazioni tra la Pol III e i  $\beta$  clamps che rendono l'enzima altamente processivo, mappano nella regione C-terminale di  $\alpha$  (ultimi 243 residui). In assenza dei  $\beta$  clamps, la replicazione da parte della Pol III è distributiva e non processiva (Lamers et al., 2006).

Esaminando la capacità replicativa di una serie di costrutti di  $\alpha$  deleti per la regione C-terminale è stato osservato che la sintesi del DNA avviene in maniera processiva, ma con minore efficienza rispetto alla sintesi condotta da  $\alpha$  *wild-type*. Quando anche il dominio OB viene rimosso (variante  $\alpha$ -929), la processività cala (Lamers et al., 2006). Pertanto anche il dominio OB coopera nel riconoscimento dei  $\beta$  clamps e quando l'oloenzima è assemblato, il dominio OB e i  $\beta$  clamps si trovano vicini. Kuryan e collaboratori hanno formulato un modello di struttura dell'oloenzima (Lamers et al., 2006). Secondo tale ipotesi, durante la replicazione  $\alpha$  lega direttamente i  $\beta$  clamps, la subunità  $\epsilon$  e  $\tau$  (figura 13).

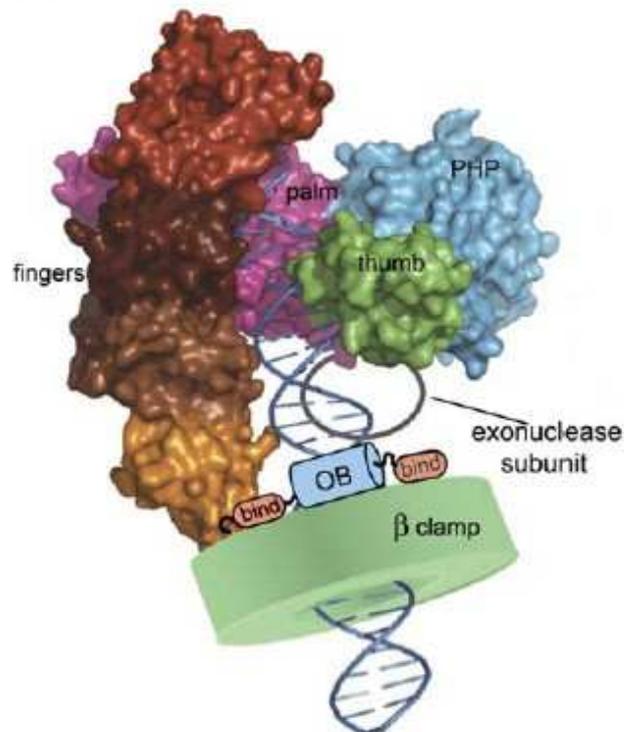


FIG 13. A: Modello di legame di  $\alpha$  ai  $\beta$  clamps e a  $\epsilon$  (exonucleasica subunità) (Lamers et al., 2006).

In particolare il legame con i  $\beta$ -clamps avverrebbe tra l'estremità del dominio Fingers e il dominio PHP, mentre l'interazione con  $\epsilon$  avverrebbe in una regione comprendente i domini PHP e Palm (residui 255-320) (Wieczorek e McHenry, 2006).

## 3.2 La subunità $\epsilon$

### 3.2.1 Funzione e struttura del gene

La subunità  $\epsilon$  della Pol III è l'enzima che svolge l'attività esonucleasica  $3' \rightarrow 5'$  ed è codificata dal gene *dnaQ* (Scheuermann et al., 1983). Il gene *dnaQ* (*mutD*) è adiacente al gene *mh*, codificante per l'RNasiH; le due ORF distano tra loro solo 64 bp e i due prodotti vengono trascritti in direzioni opposte a partire da promotori divergenti. Le sequenze per i segnali di inizio della trascrizione e della traduzione sono parzialmente sovrapposte. Si possono osservare tre Pribnow-boxes nella regione di 64 bp. Due Pribnow-boxes in tandem a 4 bp di distanza controllano l'espressione di *dnaQ* (Pq1 e Pq2), l'altra (Ph) regola *mh*. Le regioni -35 dei promotori *dnaQ* (Pq1 e Pq2) sono sovrapposte alla sequenza codificante di *mh*. Inoltre c'è una ripetizione invertita di 16 bp all'interno della regione spaziatrice di 64 bp: questa potrebbe essere una regione di controllo per l'espressione di *dnaQ* e *mh*, a cui potrebbe legarsi un effettore comune (Hisaji et al., 1983). A partire dai due promotori alternativi, il gene *dnaQ* codifica per due mRNA, *dnaQ1* e *dnaQ2*, che differiscono solo per la lunghezza dell'UTR in 5' e che sono probabilmente trascritti l'uno indipendentemente dall'altro. Il gene *mh*, invece, codifica per un unico trascritto, che risulta essere complementare alla regione 5' non tradotta di entrambi i trascritti di *dnaQ*. E' quindi possibile che si formino strutture a doppio filamento intermolecolari tra l'mRNA di *mh* e l'mRNA *dnaQ1* o *dnaQ2*. La doppia elica più corta che potenzialmente può formarsi tra *dnaQ2* e l'mRNA di *mh* è di circa 10 bp e non blocca i siti per il legame al ribosoma (RBS, Ribosome Binding Site) dei due trascritti; la più lunga, tra *dnaQ1* e l'mRNA di *mh*, è di circa 100 bp e comprende sia l'RBS che parte della regione codificante dell'mRNA di *mh*. Occorre inoltre notare che le quantità di trascritti dei due geni sono simili (Hisaji 1983, Cox e Corner, 1986). Pertanto è possibile che gli mRNA codificati da *dnaQ* e *mh* regolino l'uno l'efficienza di traduzione dell'altro, come avviene in un sistema ad RNA antisense.

### 3.2.2 La proteina DnaQ

DnaQ è una proteina di 243 aa, caratterizzata da massa molecolare pari a 27,500 Da (per una review, McHenry, 1985). L'analisi genetica di numerosi mutanti *dnaQ* (Schaaper, 1993), studi su frammenti proteolitici della subunità  $\epsilon$  ed esperimenti di two-hybrid system (Perrino et al., 1999), suggeriscono che la subunità  $\epsilon$  comprenda due domini separati da una regione cerniera detta "Q-linker". Il dominio N-terminale (residui 1-186)

è responsabile dell'attività catalitica ed è importante anche per l'interazione con la subunità  $\theta$  del complesso minimo della Pol III. Il dominio C-terminale è, invece, necessario per l'interazione con la subunità  $\alpha$ ; inoltre, sia l'interazione con  $\theta$  che con  $\alpha$  sono in grado di stimolare l'attività esonucleasica di  $\epsilon$  (Perrino et al., 1999). La subunità  $\epsilon$  possiede una bassa affinità per il DNA, probabilmente perché non è dotata di un'estesa tasca di legame per il DNA a singolo filamento. E' quindi possibile che  $\theta$  e  $\alpha$  stimolino l'attività di  $\epsilon$  *in vitro*, aumentandone l'affinità nei confronti del DNA.

L' allineamento delle sequenze amminoacidiche di diverse esonucleasi associate alle polimerasi ha rivelato regioni omologhe contenenti amminoacidi conservati, dette motivi EXO. Sono stati definiti tre motivi EXO: EXO I, II, III, ognuno contenente vari amminoacidi conservati (Blanco et al., 1992). L'importanza di tali residui, all'interno dei motivi EXO, è stata evidenziata da esperimenti di mutagenesi sito specifica in una varietà di polimerasi di diversi organismi (Blanco et al., 1992). La subunità  $\epsilon$  di *E. coli* è sprovvista del motivo EXO III (sequenza YxxxD) e presenta al suo posto un particolare motivo EXO III $\epsilon$  (sequenza HxAxxD), che condivide con l'esonucleasi della pol III di *B. subtilis* e con una serie di RNasiT (Moser et al., 1997). Il residuo D167 di EXO III $\epsilon$  di  $\epsilon$  probabilmente coordina uno ione metallico ed è particolarmente importante per la catalisi (i mutanti sono dominanti) (Schaaper, 1993). La subunità  $\epsilon$  presenta il motivo EXO I nei residui 8-21, quello EXO II nei residui 95-108, e l' EXO III nei residui 128-192 (figura 14).

La regione C-terminale (residui 187-243) di  $\epsilon$  forma un dominio indipendente da quello catalitico N-terminale. Ciò è suggerito dalla presenza di una sequenza di quattro glutammine (Q4) in posizione 194-197. Sequenze Q4 sono state infatti osservate nelle regioni linker tra i domini proteici, e si ritiene funzionino come cerniere tra domini interagenti ma distinti strutturalmente (Wootton e Drummond, 1989). La funzione del Q-linker di  $\epsilon$  potrebbe essere importante nel regolare l'attività esonucleasica e nell'orientare opportunamente i domini strutturali di  $\epsilon$  e di  $\alpha$ . Potrebbe quindi permettere una corretta interazione tra il sito attivo polimerasico e quello esonucleasico nel complesso minimo della polimerasi III (Perrino et al., 1999). Generalmente l'omologia tra  $\epsilon$  e la esonucleasi associata alla pol III di *B. subtilis* è consistente, ma termina in corrispondenza della sequenza Q4. La divergenza a valle è indicativa della diversa funzione della regione C-terminale nei due organismi: in *Escherichia coli* la subunità  $\alpha$  e la subunità  $\epsilon$  sono separate, mentre in *B. subtilis* la esonucleasi e la polimerasi sono parte dello stesso polipeptide.

La regione C-terminale è responsabile dell'associazione con  $\alpha$ . In particolare la forma troncata  $\epsilon$ 186 che presenta una delezione nella regione C-terminale, non è più in grado di legare  $\alpha$ .

La struttura di  $\theta$  complessata alla regione N-terminale di  $\epsilon$  è stata recentemente determinata (Keniry et al., 2006) e ha permesso di confermare alcune peculiarità dell'interazione tra  $\theta$  e  $\epsilon$  che erano già state descritte. Sebbene  $\theta$  non sia necessaria per l'attività esonucleasica di  $\epsilon$ , protegge  $\epsilon$ 186 dall'inattivazione termica. Inoltre

$\epsilon^{186}$  precipita irreversibilmente nelle condizioni utilizzate per studi NMR (Keniry et al., 2006), mentre il complesso  $\theta$ - $\epsilon$  è estremamente più stabile.



FIG 14: Rappresentazione schematica della proteina  $\epsilon$  della Pol III. Sono indicati i motivi conservati Exo I, Exo II e Exo III, il Q-linker e la regione C-terminale (in nero) (Taft-Benz e Schaaper, 1999).

$\epsilon$  è una proteina intrinsecamente instabile (Foster e Marinus, 1992), la cui concentrazione cellulare dipende dalla presenza dello chaperone molecolare DnaK. In ceppi che presentano mutazioni nel gene *dnaK*, il livello cellulare di  $\epsilon$  cala drasticamente (Foster e Marinus, 1992).

### 3.2.3 Mutanti *dnaQ*

Mutazioni nel gene *dnaQ* provocano una deficienza dell'attività esonucleasica, che può essere causata da due meccanismi: la mutazione può riguardare residui che influenzano l'attività catalitica, oppure può coinvolgere residui importanti strutturalmente, responsabili del legame di  $\epsilon$  alla subunità polimerasica  $\alpha$ . Ricerche iniziali evidenziarono che mutanti *dnaQ*, difettivi nell'attività 3'→5' esonucleasica dell'oloenzima Pol III, erano forti mutatori.

Si conoscono molte varianti alleliche di *dnaQ*, ed i mutanti meglio caratterizzati sono *mutD5* e *dnaQ49*. Entrambi sono forti mutatori, anche se non i più forti esistenti, e sono difettivi dell'attività esonucleasica (Taft-Benz e Schaaper, 1998). Questa deficienza è dovuta tuttavia a meccanismi diversi. Nei ceppi *mutD5* viene a mancare l'attività catalitica mentre il legame con la subunità  $\alpha$  presenta le medesime caratteristiche del *wild-type*; al contrario, le mutazioni presenti in *dnaQ49* destabilizzano la forza replicativa che, di conseguenza, si blocca in presenza di particolari strutture secondarie (Iyer, 2000). Queste conclusioni sono state ottenute mediante semplici esperimenti di complementazione, esprimendo una copia del gene *dnaQ wild type* nei due tipi di mutanti. In *dnaQ49* la subunità  $\epsilon$  *wild-type* sostituisce il suo omologo mutante (che non riesce a stabilire interazioni con la subunità  $\alpha$ ) e ripristina la normale frequenza di mutazione; questo allele è quindi definito recessivo. In *mutD5* si attua una competizione fra la subunità *wild-type* e quella deficiente nella catalisi; in questo background genetico (dominante) l'espressione della subunità  $\epsilon$  *wild type* abbassa di 5-6 volte la frequenza di mutazione (Taft-Benz e Schaaper, 1998). Nei ceppi *wild-type* di *Escherichia coli* la frequenza di

mutazione spontanea descritta equivale circa a  $10^{-10}$  per base per generazione (Drake, 1991). Nei ceppi *mutD5* si riscontrano frequenze fino a  $10^4$  volte superiori. Non è chiaro se un tale aumento del livello di mutagenesi possa essere completamente dovuto alla mancanza del proofreading. Il contributo dell'attività esonucleasica alla fedeltà della replicazione viene valutato in un fattore compreso tra  $10^2$  e  $10^3$ : in alcuni ceppi che presentano mutazioni nel locus *dnaQ* si osserva quindi una frequenza di mutagenesi superiore a quella attesa sulla base dei difetti della sola attività esonucleasica (Schaaper, 1998). Una spiegazione proposta di tale fenomeno è che il gran numero di errori nella replicazione saturi la capacità di correzione degli errori dei sistemi di riparo post replicativo amplificando di 1- 2 ordini di grandezza il fenotipo mutatore osservato (Taft-Benz e Schaaper, 1998). E' inoltre dimostrato che le interazioni fra le subunità  $\alpha$ ,  $\theta$  ed  $\epsilon$  rivestono una grande importanza e non si può escludere che nei mutanti descritti venga alterata la processività o la stabilità della Pol III (Schaaper, 1998).

Esperimenti di two o three hybrid system, hanno messo in evidenza l'effetto di  $\theta$  nei confronti di alcune varianti alleliche di *dnaQ* che presentano una ridotta interazione con  $\alpha$ . Queste varianti alleliche di *dnaQ* (*dnaQ920*, *dnaQ923*, *dnaQ924*, *dnaQ928*, *dnaQ930*, e *dnaQ932*) mostrano un fenotipo mutatore poiché presentano il dominio catalitico N-terminale mutato, un generale indebolimento strutturale di tutta la proteina e della capacità di interagire con  $\alpha$ . Pertanto, anche se i residui determinanti per l'interazione tra  $\epsilon$  ed  $\alpha$  si trovano nella regione C-terminale di  $\epsilon$ , devono esserci altri fattori coinvolti in questa interazione (Taft-Benz e Schaaper, 2004).

### 3.3 La subunità $\theta$

Il gene *holE* codifica per la subunità  $\theta$  che ha una massa molecolare di 9 kDa e che è presente nel core della Pol III. La funzione della subunità  $\theta$ , all'interno del core della Pol III, non è stata ancora completamente chiarita. La perdita di  $\theta$  ( $\Delta holE$ ) non produce un fenotipo, né si verifica un aumento nella frequenza di mutazione (Slater et al., 1994). Si può pertanto dedurre che  $\theta$  non è indispensabile per il funzionamento del core della Pol III e che la sua delezione non compromette la sintesi di DNA. Esperimenti di gel filtrazione (Studwell-Vaughan e O'Donnell, 1993) e di two hybrid system (Jonczyk et al., 1998) hanno dimostrato che esiste una stretta interazione tra  $\theta$  ed  $\epsilon$ , ma non tra  $\theta$  ed  $\alpha$ .

Inoltre,  $\theta$  è in grado di aumentare di circa 2.5 volte l'attività esonucleasica  $3' \rightarrow 5'$  di  $\epsilon$  (Studwell-Vaughan e O'Donnell, 1993) producendo un effetto antimutatore. Pertanto  $\theta$ , pur non essendo essenziale, potrebbe svolgere indirettamente un ruolo chiave nella replicazione del DNA attraverso l'interazione con la subunità  $\epsilon$ . Da studi recenti (Taft-Benz e Schaaper, 2004) è emerso che  $\theta$  potrebbe essere un fattore stabilizzante per  $\epsilon$  che è intrinsecamente instabile (Foster e Marinus, 1992). Di conseguenza, da un punto di vista evolutivo, la frequenza di mutazione potrebbe essere ridotta grazie al modesto contributo di  $\theta$  (Taft-Benz

e Schaaper, 2004). Nei complessi replicativi di batteri gram-negativi l'enzima polimerasi e l'esonucleasi sono due proteine separate. Negli altri organismi nei quali la polimerasi e la esonucleasi fanno parte dello stesso polipeptide, non sono state trovate proteine, o loro parti, omologhe a  $\theta$ . Questo fa pensare che  $\theta$  possa essersi evoluta per proteggere la subunità  $\epsilon$  (Taft-Benz e Shaaper, 2004). Oltre a stabilizzare  $\epsilon$  nel core della Pol III,  $\theta$  potrebbe sequestrare la forma libera di  $\epsilon$  nella cellula e bloccarla nel core dell'oloenzima; in questo modo  $\epsilon$  risulterebbe molto più stabile rispetto alla forma non complessata (De Rose et al., 2003). Non è ancora chiaro come avvenga l'assemblaggio del core della Pol III, ma la formazione del complesso di  $\epsilon$  con  $\theta$  potrebbe assicurare un quantitativo di subunità  $\epsilon$  maggiore proprio grazie alla stabilizzazione da parte di  $\theta$  (Taft-Benz e Schaaper, 2004).

Recentemente è stata identificata una proteina del batteriofago P1 che è in grado di sostituire  $\theta$  nel core della Pol III e, proprio per questa caratteristica funzionale è stata chiamata HOT (Homolog Of Theta) (Chikova e Shaaper, 2005). La replicazione del fago P1 dipende quasi interamente dal sistema replicativo dell'ospite. Essendo  $\epsilon$  instabile, HOT potrebbe fungere da agente stabilizzante per fornire un beneficio al fago che potrebbe avere a disposizione una maggiore quantità di core e, in ultima analisi, anche di oloenzima (Chikova e Shaaper, 2005).

## 4. Stabilità delle proteine

### 4.1 Proteasi batteriche

L'espressione genica è regolata a livello della trascrizione, della traduzione e dei processi post-traduzionali. Tra quest'ultimo gruppo di meccanismi di regolazione è inclusa la proteolisi che è richiesta non solo per controllare l'emivita di un polipeptide, ma anche per agire in modo specifico nei confronti delle proteine danneggiate che, se non vengono idrolizzate, potrebbero funzionare in maniera alterata. Una proteina può risultare danneggiata in seguito ad errori di sintesi oppure per la perdita di struttura terziaria in seguito a cambiamenti fisiologici o ambientali. Pertanto le proteasi svolgono un ruolo chiave nel turnover delle proteine con un errato ripiegamento ("misfolded"), danneggiate o che hanno perso la corretta struttura ("unfolded"). Le proteasi utilizzano tali proteine come substrato per effettuare una "degradazione regolatoria" di tutte le proteine la cui funzione deve essere strettamente controllata, come ad esempio i fattori di trascrizione.

Fino ad ora poche ricerche si sono occupate dell'interazione tra la proteolisi, il mantenimento dell'integrità genica e la mutagenesi (Al Mamun e Humayun, 2008). Concretamente si può supporre che il fallimento della proteolisi possa influire sulla fedeltà della replicazione del DNA. Se il meccanismo di degradazione delle proteine coinvolte nella replicazione, riparazione o ricombinazione del DNA fosse difettoso, potrebbero

verificarsi errori di replicazione, di correzione o di riparazione del DNA e pertanto un problema di mantenimento della stabilità genetica.

La degradazione intracellulare delle proteine gioca un ruolo fondamentale nella maggior parte dei processi fisiologici come la rimozione di polipeptidi danneggiati o di proteine provviste di una specifica sequenza segnale. La rimozione selettiva di proteine regolatorie come i fattori di trascrizione o quelle coinvolte nella trasduzione del segnale rappresenta una strategia efficiente e rapida per controllare i punti fondamentali di molti processi cellulari come la crescita cellulare, la divisione e la morte programmata.

Nei procarioti, sebbene un grande numero di proteasi sia conosciuto, cinque sono considerate le più importanti. Quattro di queste sono proteasi citosoliche: Lon, ClpXP, ClpAP e HslVU (ClpQY), mentre la quinta è HflB (FtsH) una proteasi ancorata alla membrana (per una review, Gottesman, 2003). La maggior parte della proteolisi intracellulare è realizzata da proteasi che funzionano a spese di idrolisi di ATP. Il dominio ATPasico di queste proteasi media il riconoscimento del substrato; gli elementi di riconoscimento non sono ben definiti e non sono, probabilmente, sequenze aminoacidiche specifiche (Gottesman, 1996).

HflB è l'unica proteasi essenziale per *E. coli*. La perdita della proteasi Clp (ClpXP/AP) determina invece un aumento della mutagenesi e in particolare della frequenza di mutanti che presentano uno spostamento del modulo di lettura (mutazioni frameshift). L'effetto mutatore è assente in cellule che presentano una delezione del gene *dinB*, suggerendo che l'effetto mutatore è mediato dalla Pol IV (Al Mamun e Humayun 2008).

La proteolisi citosolica nei procarioti di proteine "misfolded" o di proteine specificatamente da degradare è mediata da un complesso proteico simile al proteasoma delle cellule eucariotiche (Mogk et al., 2007). Questo consiste di un dominio ATPasico (ClpA, ClpX) e di un componente proteolitico legato covalentemente (esempio, Lon) o diffusibile (esempio, ClpP) (per una review, Sauer et al., 2004).

Nonostante la peptidasi ClpP di *Escherichia coli* non presenti omologia di sequenza con le proteine del proteasoma eucariotico, forma comunque una struttura simile. Consiste di due anelli eptamerici che formano un core a forma di barile con il sito attivo proteolitico posto all'interno di questa struttura. L'accesso a questo sito è controllato da stretti domini che non permettono il passaggio di proteine con una corretta struttura terziaria. La traslocazione verso il sito attivo proteolitico è mediata da ClpA che è associata al core di ClpP. ClpA fa parte della famiglia di proteine AAA+ (figura 15).

Sulla base dell'omologia di sequenza, le proteasi ATP-dipendenti vengono raggruppate nella famiglia delle proteine AAA+ (ATPase Associated with cellular Activities). Le proteine AAA+ sono coinvolte in una grande varietà di processi cellulari quali la replicazione, la trascrizione e la proteolisi. Queste proteine contengono un dominio ATPasico e presentano residui al C-terminale in grado di riconoscere la presenza di un  $\gamma$ -fosfato in un nucleotide e di catalizzarne l'idrolisi. La regione C-terminale contiene anche una regione che interagisce con il substrato detto dominio SSD (Substrate Sensor and Discrimination) (Smith et al., 2001). Come altre proteine AAA+, le proteasi ATP dipendenti sono enzimi oligomerici che hanno strutture simili ad anello o simili ad un

barile con una cavità centrale che è detta “camera proteolitica” ed è il sito nel quale la proteina bersaglio viene traslocata e subisce proteolisi. La degradazione dei substrati proteici richiede la presenza di ATP che viene idrolizzato.

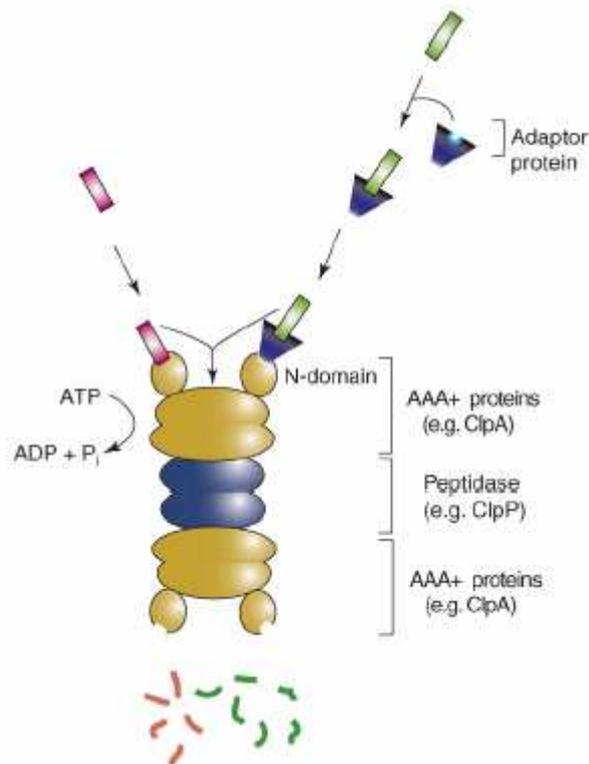


FIG 15: Componenti del proteasoma procariotico; l'attività proteolitica dipende anche dalla idrolisi di ATP (Mogk et al., 2007).

Le proteasi ATP- dipendenti possono essere divise in due gruppi: fanno parte del primo gruppo i complessi che contengono in subunità diverse l'ATPasi e la proteasi e del secondo quei complessi che contengono nella stessa catena polipeptidica i due domini funzionali. Fanno parte del primo gruppo le proteasi come Clp, mentre del secondo le proteasi come Lon; nonostante la differenza nell'organizzazione delle subunità, la struttura e i meccanismi sono simili. In particolare è emerso da studi cristallografici che il dominio ATPasico serve per destrutturare, traslocare e attivare il sito proteolitico (Baker e Sauer, 2006). In presenza di ATP, le proteasi Clp e Lon degradano il substrato proteico senza generare prodotti intermedi ma solamente proteine non digerite o peptidi totalmente digeriti (Lee et al., 2001), elemento che suggerisce che la proteolisi è processiva. Dal momento che si verificano diversi cicli di legame dell'ATP e di idrolisi è plausibile pensare che anche l'attività ATPasica contribuisca a rendere la proteolisi processiva.

Sebbene la specificità che le proteasi ATP- dipendenti mostrano per il substrato non sia ben definita, è stato osservato che la traslocazione inizia in seguito all'interazione tra proteina substrato e il dominio ATPasico (Licht e Lee, 2008). Nel caso della proteasi Lon non è stata identificata una sequenza consenso (Nishii et al.,

2002) mentre nel caso della proteasi Clp è stata individuata una sequenza di 11 aminoacidi detta *ssrA* (Gottesman et al., 1998) e altre corte sequenze che fungono da segnale di riconoscimento.

La presenza di ATP induce una serie di cambiamenti conformazionali nel dominio della proteasi, determinante per la destrutturazione della proteina bersaglio. Il passaggio successivo è la traslocazione della proteina substrato nella cavità della proteasi. La destrutturazione e la traslocazione della proteina bersaglio sono considerati i passaggi che limitano la velocità di reazione (Kenniston et al., 2003) e che sono realizzati attraverso cicli ripetuti di idrolisi dell'ATP che induce questi cambiamenti conformazionali.

## 4.2 Regola dell'N-terminale

La regola dell'N-terminale stabilisce che la emivita di una proteina è determinata dalla natura dei suoi residui N-terminali. Questo principio fondamentale che regola la proteolisi è conservato sia nei batteri che nei mammiferi (per una review, Mogk et al., 2007) poiché sebbene i procarioti e gli eucarioti impieghino un diverso meccanismo proteolitico per degradare i substrati sulla base della regola dell' N-terminale, condividono comunque i principi per il riconoscimento dei bersagli proteici.

La degradazione delle proteine dipende dalla presenza di residui aminoacidi stabilizzanti o destabilizzanti. Nel 1986, Varshavsky e collaboratori scoprirono che diversi costrutti della proteina  $\beta$ -galattosidasi di *Escherichia coli* prodotti in *Saccharomyces cerevisiae*, esibivano una emivita diversa che andava da 20 ore a meno di 3 minuti (Bachmair et al., 1986). La stabilità della proteina ingegnerizzata dipendeva dalla natura degli aminoacidi N-terminali. Sulla base di questi e di altri esperimenti gli aminoacidi sono stati classificati come residui stabilizzanti o destabilizzanti. Questi segnali di degradazione sono stati esaminati in vari organismi tra cui *E. coli* (Tobias et al., 1991) e sono stati definiti "N-degron" (figura 16). I residui destabilizzanti sono rappresentati dagli aminoacidi aromatici (fenilalanina, triptofano e tirosina) e dalla leucina. La presenza o la generazione di questi "N-degron" causa proteolisi nella proteina.

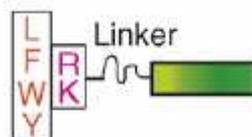


FIG 16: I residui idrofobici come Leu, Phe, Trp, Tyr e i residui basici rappresentano rispettivamente residui destabilizzanti primari e secondari e contribuiscono alla definizione di "N- degrons". Il substrato soggetto a proteolisi per la regola dell' N-terminale contiene una regione flessibile (linker) che separa la regione N-terminale dal substrato. Tale linker è richiesto per il trasferimento e la degradazione da parte delle proteasi ClpA/ClpP (Mogk et al., 2007).

Tutte le proteine neo-sintetizzate contengono all' N-terminale una metionina (formil-metionina nei procarioti) che è un residuo stabilizzante secondo la regola dell'N-terminale. Pertanto un N-degron potrebbe essere

prodotto a partire da un “pre- N-degron”. Se la posizione 2 della catena proteica è occupata da un aminoacido con una piccola catena laterale come la glicina, la cisteina, la serina o l’alanina, residui stabilizzanti secondo la regola dell’N-terminale (Giglione et al., 2004), la metionina N-terminale viene rimossa dalle proteine neosintetizzate ad opera della metionina amminopeptidasi (MetAp). La specificità della MetAp, conservata nei procarioti e negli eucarioti, indica che c’è stata una coevoluzione nei meccanismi connessi alla regola dell’N-terminale. Pertanto la rimozione di residui stabilizzanti all’N-terminale e/o la generazione di nuovi residui destabilizzanti funzionano come punto di entrata di proteasi la cui azione è influenzata dai residui presenti all’N-terminale. Ad esempio un N-degrone potrebbe trovarsi all’interno di una proteina ed essere accessibile solo in seguito ad un taglio proteolitico. Attraverso l’aggiunta o la modificazione di aminoacidi è possibile generare residui destabilizzanti all’N-terminale che possano poi essere processati da proteasi specifiche.

Secondo la regola dell’N-terminale esistono non solo residui destabilizzanti fino ad ora descritti (detti anche destabilizzanti primari), ma anche residui destabilizzanti secondari. In *Escherichia coli*, per esempio, l’arginina e la lisina sono considerati residui destabilizzanti secondari perché reclutano residui destabilizzanti primari come la leucina o la fenilalanina attraverso l’azione enzimatica di una leucil/fenilalanil-tRNA transferasi (Tobias et al., 1991) (figura 17).

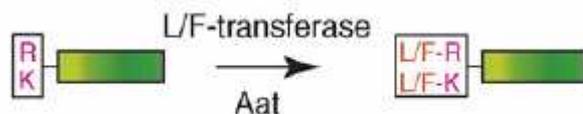


FIG 17: La reazione enzimatica catalizzata dalla leucil/fenilalanil-tRNA transferasi (L/F- transferase) converte i residui destabilizzanti terziari e secondari in primari.(Mogk et al., 2007).

La proteina ClpS funge da adattatore per le proteasi ClpA/ ClpP (Lupas e Karetke, 2003) poiché possiede una regione ricca in residui acidi e idrofobici che funziona come superficie complementare nell’interazione con l’ “N-degron”. ClpS interagisce direttamente con l’estremità N-terminale di una proteina che deve essere degradata e che contiene, secondo la regola dell’estremità N-terminale, degli aminoacidi destabilizzanti. Infatti, mutanti di *E. coli* deficienti per ClpA o per la proteina adattatrice ClpS, non sono in grado di degradare proteine con residui destabilizzanti all’ N-terminale (Erbse et al., 2006). ClpS si lega a proteine che contengono all’N-terminale i residui Phe, Tyr, Trp, Leu. Il substrato è trasferito efficientemente da ClpS a ClpA solamente se è presente una regione non strutturata tra l’ ”N-degron” e la restante parte della proteina normalmente strutturata. Questa caratteristica riscontrata nelle proteine degradate da ClpA rende il meccanismo degradativo ancora più selettivo e stringente.

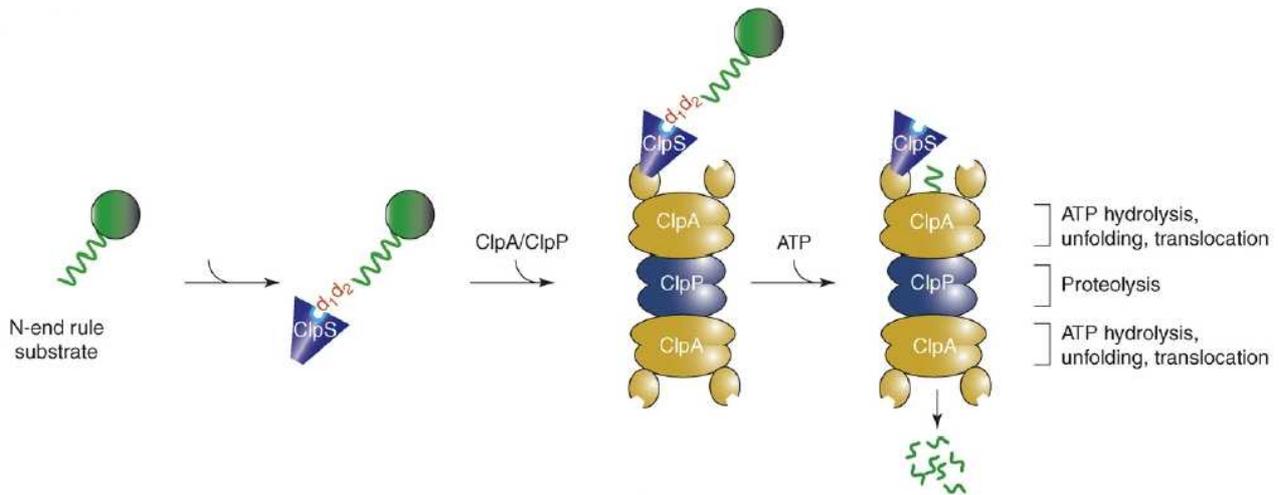


FIG 18: Gli aminoacidi idrofobici (Phe, Leu, Trp, Tyr) funzionano come residui destabilizzanti ( $d_1$ ) e i residui basici (Arg, Lys) come residui destabilizzanti secondari ( $d_2$ ). La proteina adattatore ClpS lega direttamente le proteine bersaglio che contengono i residui destabilizzanti e poi interagisce con l'estremità N-terminale di ClpA (Mogk et al., 2007).

### 4.3 Riconoscimento di specifici residui nella regione C-terminale

Sauer e collaboratori (McGinness et al., 2006) si accorsero che la stabilità delle proteine può essere influenzata dai residui C-terminali. In particolare, le sequenze ricche in aminoacidi non polari causano una rapida degradazione. Le proteine nelle quali gli ultimi cinque aminoacidi sono polari o carichi non vengono degradati. Questo suggerisce che nella cellula esistano delle proteasi capaci di riconoscere e iniziare la degradazione a partire da questi segnali. La proteina Tsp (tail-specific protease), proteina periplasmatica, potrebbe riconoscere e degradare queste proteine instabili a livello della loro regione C-terminale nel periplasma o nella membrana plasmatica, mentre nel citoplasma sarebbe ClpXP la proteasi responsabile del riconoscimento dei residui non polari (Gottesman, 1996).

### 4.4 Gli chaperones molecolari

La corretta struttura tridimensionale di una proteina (conformazione nativa), sia che si tratti di un enzima o di un elemento strutturale, è essenziale perché la sua funzione biologica possa essere svolta. Gli chaperones molecolari sono preposti al mantenimento nella cellula di proteine di altamente funzionali.

Gli chaperones molecolari sono una classe funzionale di molecole che transientemente agiscono in vivo per assicurare il corretto ripiegamento ed assemblaggio di proteine cellulari (Ellis, 1993). Gli chaperones molecolari contribuiscono al giusto ripiegamento di proteine appena sintetizzate, alla loro associazione in strutture più complesse, al trasporto attraverso la membrana cellulare ed infine cooperano nella riparazione di polipeptidi danneggiati da uno stress. Gli chaperones molecolari permettono per esempio di limitare i danni

che le proteine potrebbero subire in caso di shock termico. Il fallimento del corretto ripiegamento di una proteina può essere deleterio non solo per la perdita della sua funzionalità, ma anche per la tossicità che gli aggregati proteici possono avere per la cellula (Saibil, 2008).

Gli chaperones molecolari sono debolmente espressi in condizioni ottimali di crescita e sono drammaticamente sovraespressi in presenza di uno shock termico o in presenza di un altro tipo di stress che determina un aumento del numero di proteine che presentano una conformazione errata. Per questo motivo molti chaperones molecolari sono stati classificati come proteine indotte in seguito a shock termico (Heat Shock proteins, Hsp) o indotte in seguito a stress (Stress responsive proteins).

Gli chaperones molecolari riescono a legare (funzione holder) un bersaglio proteico che si trova in forma non nativa (ovvero che presenta una struttura tridimensionale scorretta) per prevenire l'aggregazione irreversibile e rilasciare la proteina quando questa ha riacquisito la corretta conformazione nativa (funzione folder).

#### **4.4.1 Gro-EL, Gro-ES**

Il fattore sigma RpoH ( $\sigma^{32}$ ) induce l'espressione di un insieme di geni in caso di shock termico o in caso di presenza di proteine non ripiegate nella corretta struttura tridimensionale. RpoH favorisce la trascrizione di Gro-E anche in altre condizioni di stress quali i danni al DNA, stress ossidativo, presenza di antibiotici e di metalli pesanti, infezione fagica, assenza di fonti di carbonio o di aminoacidi nella cellula (Morita et al., 2000). L'operone groEL/ES, che codifica per lo chaperone molecolare GroE fa parte dunque di questo pool di geni considerati in grado di rispondere alle condizioni di stress per la cellula. GroE è necessaria, indipendentemente dalla temperatura alla quale si trova la cellula di *E. coli*, per favorire il giusto ripiegamento delle proteine essenziali e per permettere il mantenimento della giusta conformazione. Il livello cellulare di Pol V dipende da GroE dal momento che lo chaperone molecolare interagisce con UmuC, la subunità polimerasica della Pol V, proteggendola dalla degradazione (Donnelly e Walker, 1992).

GroEL è un tetradecamero a forma di doppio anello (figura 19) che funziona con GroES, partner con il quale forma un complesso molecolare funzionale. La proteina bersaglio di GroEL viene intrappolata nella cavità centrale del tetradecamero e, in presenza di GroES e di ATP idrolizzato, potrà avvenire il corretto ripiegamento. GroEL è in grado di catturare un grande numero di proteine (Viitaneen et al., 1992).

La proteina in forma non-nativa è catturata attraverso il legame a tre o quattro dei sette siti idrofobici presenti nell'anello aperto di GroEL (Farr et al., 2000). Il legame dell'ATP e di GroES a GroEL inducono cambiamenti conformazionali tali da intrappolare il substrato proteico non nativo in un ambiente idrofilico nel quale può avvenire il corretto ripiegamento (Horwich et al., 2006).

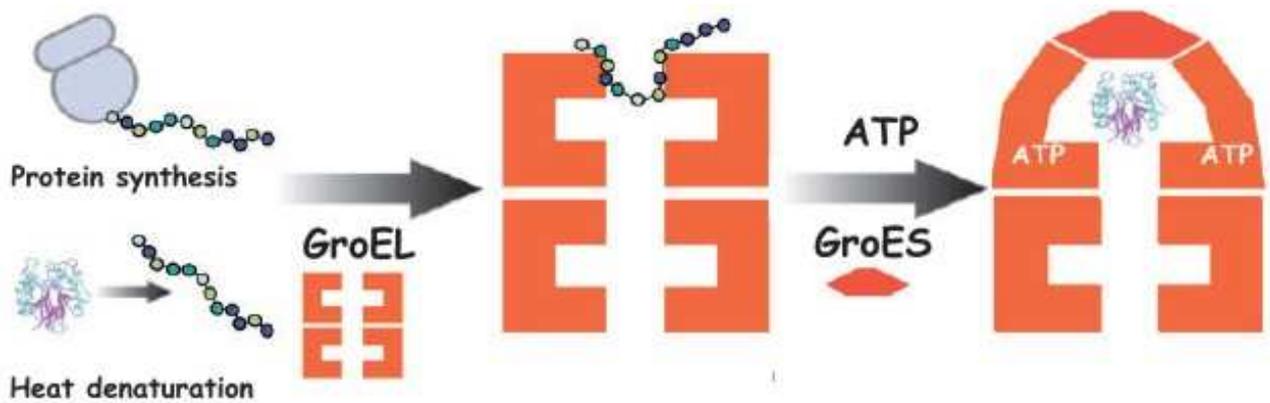


FIG 19: Disegno schematico che mostra come lo chaperone molecolare GroEL funziona. Tale chaperone molecolare coadiuva la sintesi proteica proteggendo le proteine neosintetizzate a spese di ATP (Taguchi, 2005).

#### 4.5 Keio collection e ceppi deleti per proteasi e chaperones molecolari

La disponibilità crescente di nuove sequenze genetiche ha permesso di allargare le conoscenze, a livello molecolare, degli esseri viventi. *E. coli* K12 è stato uno degli organismi meglio studiati in biologia molecolare e, nonostante questo, molti dei suoi geni funzionali non sono stati ancora caratterizzati.

Una risorsa per effettuare analisi sistematiche di geni con funzione sconosciuta è rappresentata dall'utilizzo della collezione Keio, che è composta da 3985 mutanti deleti (Baba, 2006) in singoli geni non essenziali di *E. coli* K12. Le regioni codificanti per geni non essenziali sono state sostituite da una cassetta contenente kanamicina.

### 5. Applicazioni: costruzione di mutatori

#### 5.1 Ceppi mutatori

Un organismo si definisce mutatore quando presenta una maggiore frequenza di mutazioni spontanee a carico del proprio genoma rispetto al corrispondente *wild type*. Un fenotipo simile può essere il risultato di diverse varianti genotipiche (Miller, 1998). I meccanismi che, una volta alterati, determinano frequenze di mutazione superiori alla norma sono: l'apparato replicativo (I), il riparo ossidativo (II), i sistemi di escissione di basi alterate o danneggiate (III), i sistemi di mismatch repair (IV), alcuni processi ricombinativi (V). A seconda dell'importanza dei sistemi alterati e della coesistenza di più loci mutati, la frequenza di mutazione può essere superiore da poche unità fino a 10.000 volte rispetto al *wild type*. I ceppi mutatori costituiscono un importante

strumento di studio poiché la loro caratterizzazione permette di indagare i meccanismi di mutagenesi e riparo del DNA, il ruolo dei loci che conferiscono fenotipi mutatori in alcune malattie umane e nel cancro, e di valutare l'importanza degli organismi mutatori nel corso dell'evoluzione (Miller, 1998).

In tabella sono elencati alcuni fra i più noti e meglio caratterizzati ceppi mutatori di *Escherichia coli*. Differiscono tra loro per il locus interessato, il tipo di mutazione che si verifica più frequentemente e la frequenza di mutazione:

<b>Mutazione</b>	<b>Specificità</b>	<b>Frequenza</b>
mutT	Trasversioni	Moderata
mutH	Transizioni	Alta
mutS	Transizioni; frameshift	Alta
mutY	Trasversioni	Alta
mutD	Tutti i tipi; frameshift	Molto alta

Una delle cause primarie che porta ad una elevata frequenza di mutazioni spontanee è una replicazione poco fedele che porta all'insorgenza di errori durante la sintesi del DNA. Il genotipo mutatore *mutD5* è un esempio di questo tipo.

## 5.2 Produzione ed impiego di ceppi mutatori

Sono state sviluppate numerose tecniche per isolare i ceppi mutatori, soprattutto di *Escherichia coli* e di *Salmonella*, basate su analisi fenotipiche quali l'insorgenza della resistenza ad antibiotici e la reversione da auxotrofia a prototrofia per un determinato nutriente.

In una popolazione batterica, mutanti spontanei si formano frequentemente e alcuni di questi sono mutatori. Il vantaggio che essi offrono è l'enorme adattabilità all'ambiente circostante dal momento che mutano molto più velocemente della restante parte della popolazione. Il mantenimento della popolazione di mutatori dipende dal risultato della competizione con i non mutatori (Arjan e De Visser, 2002). In *E. coli* coltivato in terreno liquido per circa 30 generazioni, la frequenza di mutatori è intorno a  $10^{-5}$ , mentre in *Salmonella typhimurium* è di circa  $10^{-6}$  (Le Clerc, 1998). L'utilizzo di agenti mutageni può incrementare la proporzione di mutatori nella popolazione fino a  $10^{-3}$  (Mao, 1997), mentre selezioni successive, ottenute trasferendo le cellule in diversi terreni, possono portare la frequenza di mutatori al 100%. Per esempio, è stato progettato un singolo terreno che seleziona vari fenotipi in successione (reversione da auxotrofia a prototrofia), permettendo di ottenere fino al 100 % di colonie contenenti mutatori (Miller et al., 1999).

I ceppi mutatori possono essere utilizzati per la mutagenesi casuale di particolari geni allo scopo di ottenere proteine con caratteristiche nuove. L'ingegneria proteica si occupa della stabilizzazione e della modificazione funzionale di enzimi. Un metodo utile in grado di introdurre mutazioni in geni bersaglio è detto "evoluzione orientata di enzimi", e consiste in un approccio di tipo evolutivistico che utilizza gli stessi strumenti dell'evoluzione naturale, la mutazione e la ricombinazione. Utilizzando tale approccio, è possibile ridurre i tempi necessari per ottenere una particolare variante proteica. Dal punto di vista sperimentale, questo si traduce nella costruzione di librerie di proteine mutanti. L'evoluzione è guidata perché si decidono in partenza le caratteristiche che si vogliono sviluppare e, a tal scopo, si utilizzano opportuni test di screening o metodiche di selezione. Ad ogni ciclo di mutazione vengono valutate le proprietà dei mutanti ottenuti; quelli che dimostrano un miglioramento verso l'attività desiderata, sono ulteriormente sottoposti ad un passaggio di mutagenesi, generando così una nuova libreria. Una tecnica *in vivo* per ottenere librerie di mutanti, si basa proprio sull'uso di ceppi mutatori. Questo significa che un vettore plasmidico che si replica all'interno di queste cellule accumula un certo numero di mutazioni, modulabile in funzione del numero di generazioni di permanenza.

I ceppi mutatori possono essere usati anche per evolvere nuove caratteristiche fenotipiche. L'uso di microrganismi *wild type* nelle biotrasformazioni è frequentemente vincolato dalla loro limitata tolleranza ad alte concentrazioni di metaboliti o solventi organici, spesso indispensabili nelle produzioni industriali. In molti casi, ceppi con caratteristiche più adatte per l'utilizzo tecnologico sono stati generati tramite mutagenesi casuale, seguita da selezione o screening. Le mutazioni spontanee possono portare a diversità in una popolazione sufficientemente ampia; tuttavia, ulteriori mutazioni vengono spesso introdotte usando radiazioni ultraviolette o mutageni chimici. Sfortunatamente, la maggior parte di questi agenti mutageni sono difficilmente modulabili, comportando così frequentemente la morte delle stesse cellule trattate. I ceppi mutatori possono, invece, eliminare il bisogno di aumentare la frequenza di mutazione con agenti mutageni fisici o chimici, ma la loro applicazione è stata finora limitata dalla propria instabilità genetica e dalla necessità di stabilizzare i ceppi dopo la selezione (Selifonova et al, 2001).

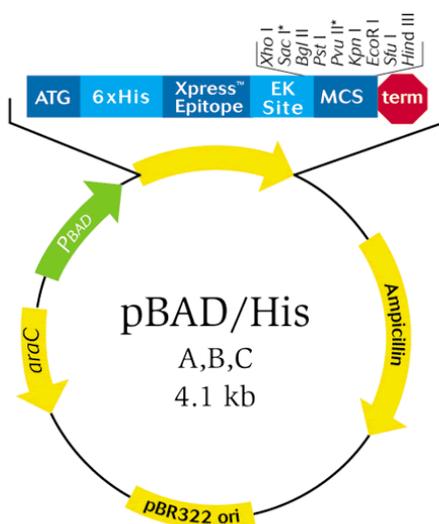
# MATERIALI E METODI

## 1. Ceppi batterici e vettori di espressione

Nell'ambito di questa tesi sono stati utilizzati i seguenti ceppi di *Escherichia coli*:

CEPPO	GENOTIPO	PROVENIENZA
BW25113	$\Delta(\text{araD-araB})567$ , $\Delta\text{lacZ4787}>::\text{rmB-3}$ , $\lambda$ , $\text{rph-1}$ , $\Delta(\text{rhaD-rhaB})568$ , $\text{hsdR514}$	CGSC
JW0429-1	BW25113 $\Delta\text{lon-725}>::\text{kan}$	CGSC, Keio collection
JW0427-1	BW25113 $\Delta\text{clpP723}>::\text{kan}$	CGSC, Keio collection
JW3903-5	BW25113 $\Delta\text{hslV720}>::\text{kan}$	CGSC, Keio collection
JW0013-3	BW25113 $\Delta\text{dnaK734}>::\text{kan}$	CGSC, Keio collection
JW4103-1	BW25113 $\Delta\text{groL768}>::\text{kan}$	CGSC, Keio collection
JW0462-1	BW25113 $\Delta\text{htpG757}>::\text{kan}$	CGSC, Keio collection
JW0426-1	BW25113 $\Delta\text{tig-722}>::\text{kan}$	CGSC, Keio collection
BW11334	$\Delta(\text{argF-lac})169$ , $\text{recA1907}>::\text{cat-aadA}$ , $\text{thi-1}$ , $\Delta(\text{phnP-phnD})3330$ , $\text{creC510}$	CGSC
TOP10	$\text{mcrA}$ , $\Delta(\text{mrr-hsdMRS-mcrBC})$ , $\Delta\text{lacX74}$ , $\phi 80\text{lacZ}\Delta\text{M15}$ , $\text{deoR}$ , $\text{endA1}$ , $\text{recA1}$ , $\text{araD139}$ , $\Delta(\text{araA-leu})7697$ , $\text{galU}$ , $\text{galk}$ , $\text{rpsL}$ , $\text{nupG}$	Invitrogen
BW25993	$\Delta(\text{araD-araB})567$ , $\text{rph-1}$ , $\Delta(\text{rhaD-rhaB})568$ , $\text{hsdR514}$	CGSC

Il vettore d'espressione utilizzato in questo lavoro è il pBAD/HisB (Invitrogen), derivato da plasmidi di tipo ColE1.



\* Frame-dependent variations

L'espressione è sotto il controllo del promotore dell'operon *ara* (arabinosio), ed è regolata dalla presenza della proteina AraC, espressa costitutivamente dallo stesso vettore. Per l'espressione di *dnaQ*, *dnaE* e delle forme tronche di questi geni è stato aggiunto al terreno di coltura arabinosio come induttore; le concentrazioni utilizzate saranno indicate nei paragrafi successivi.

## 1.1 Terreni di coltura

I ceppi batterici utilizzati sono stati coltivati in appositi terreni liquidi. Per la coltivazione in piastra, il terreno è stato addizionato di agar 1.5%.

Terreno	Composizione
LB	10 g/L di triptone 5 g/L di estratto di lievito 10 g/L di NaCl
SOC	10 g/L di triptone 5 g/L di estratto di lievito 10 g/L di NaCl MgCl <sub>2</sub> 10 mM, KCl 2,5 mM glucosio 2 g/L

Qualora necessari, gli antibiotici sono stati aggiunti alle seguenti concentrazioni ai terreni di coltura liquidi o solidi:

Antibiotico	Concentrazione finale
Kanamicina	40 µg ml <sup>-1</sup>
Ampicillina	100 µg ml <sup>-1</sup>
Tetraciclina	15 µg ml <sup>-1</sup>
Cloramfenicolo	20 µg ml <sup>-1</sup>
Rifampicina	100 µg ml <sup>-1</sup>
Streptomicina	30 µg ml <sup>-1</sup>

## 1.2 Curve di crescita

Le curve di crescita dei ceppi di *E. coli* contenenti i vari costrutti sono state ottenute spettrofotometricamente determinando l'assorbanza a 600 nm delle colture allevate in beute contenenti 25 mL di LB, mantenute in agitazione e a temperatura costante (37 °C).

## 2. Metodiche generali

### 2.1 Estrazione di DNA genomico

L'estrazione del DNA genomico è stata effettuata da una coltura overnight di cellule batteriche tramite il kit Dneasy Blood & Tissue Kit (Qiagen). Il DNA estratto è stato risospeso in 100 µl di acqua sterile e quantificato determinando l'assorbanza a 260 nm.

### 2.2 PCR (Polymerase Chain Reaction)

Ogni reazione di amplificazione è stata condotta in un volume finale di 50 µl in presenza di:

MgCl<sub>2</sub>      2 mM,  
dNTP mix 0.2 mM (concentrazione finale di ciascun nucleotide)  
primers    0.2 µM (concentrazione finale di ogni primer <sup>1</sup>)

L'enzima e il suo specifico tampone impiegati sono stati scelti in base alle caratteristiche della sequenza da amplificare e al successivo utilizzo del frammento amplificato. Pertanto gli enzimi, le temperature (di annealing e di extension) e il numero dei cicli di reazione sono indicati nei paragrafi successivi.

Tutte le reazioni di PCR sono state realizzate con il termociclatore Teche TC-3000.

### 2.3 Elettroforesi in gel d'agarosio

---

<sup>1</sup> La sintesi dei primers utilizzati nel presente lavoro è stata affidata alla ditta Sigma Genosys

Per la preparazione di tutti i gel analitici è stato utilizzato Certified™ Molecular Biology Agarose (Bio-Rad) e tampone tris-acetato (TAE, 1X) (Fritsch et al., 1989) ad una concentrazione 1% (peso / volume).

Ai campioni di DNA da caricare in gel è stato aggiunto l'addensante 6 × DNA Loading Dye (Fermentas), la corsa è stata condotta in camera elettroforetica (Bio-Rad) impiegando il tampone TAE (1X).

Dopo la corsa, il gel è stato incubato in agitazione per 30' a temperatura ambiente in una soluzione contenente Etidio Bromuro (1µg/ml) in TAE (1X) e le bande sono state visualizzate con Molecular Imager Gel Doc (Bio-Rad).

## 2.4 Recupero e purificazione di DNA da gel di agarosio

Il recupero selettivo da gel dei frammenti di DNA di interesse, amplificati attraverso PCR, è stata effettuata utilizzando il kit Freeze and Squeeze (Bio-Rad), seguendo le indicazioni fornite dalla ditta. Il DNA eluito è stato concentrato con filtri Montage-PCR (Millipore) e risospeso in un opportuno volume di acqua sterile milli-Q. Il DNA è stato poi quantificato attraverso elettroforesi (§ 2.3).

## 2.5 Clonaggio genico

Il vettore pBAD e l'insero da clonare sono stati digeriti con gli opportuni enzimi di restrizione (NEB). Per prevenire la propria autosaldatura, il vettore è stato sottoposto a trattamento di defosforilazione a livello delle estremità 5' per 30 minuti a 37 °C con 0.5 unità di Calf Intestine Alkaline Phosphatase (CIAP, Fermentas).

Vettore e insero sono stati purificati mediante l'aggiunta di un uguale volume di una soluzione contenente fenolo e cloroformio, quindi precipitati in 0.1 volumi di sodio acetato 3 M pH 5.5 e 2.5 volumi di etanolo assoluto. Dopo un' incubazione overnight a -80 °C e una centrifugazione a 14000 g per 30', (Eppendorf Centrifuge 5415 R) i pellets ottenuti sono stati lavati con etanolo 70% e, dopo asciugatura, sospesi in un opportuno volume di acqua sterile, quindi quantificati attraverso elettroforesi (§ 2.3).

La reazione di ligazione tra insero e vettore è stata effettuata utilizzando 2.5 unità dell'enzima DNA ligasi T4 (NEB).

La quantità di insero (ng) da utilizzare è stata calcolata utilizzando la seguente formula:

$$[\text{ng vettore} \times \text{dimensioni (kb) insero} \times \text{rapporto molare (insero: vettore)}] / \text{dimensioni (kb) vettore}$$

Sono stati utilizzati 50 ng di vettore per reazione e un rapporto molare (insero: vettore) pari a 3:1.

La reazione è stata effettuata in un volume finale di 10 µl, a 16 °C overnight.

## 2.6 Preparazione di cellule elettrocompetenti

Colonie singole di *Escherichia coli* sono state trasferite dalle piastre Petri contenenti terreno solido, in beute contenenti 10 ml di terreno LB addizionato con gli opportuni antibiotici. Le beute sono state incubate overnight a 37 °C in agitazione continua (180 rpm). Le colture sono state quindi diluite 1:100 in 50 ml dello stesso terreno ed incubate 3 ore in modo da ottenere una densità di popolazione pari ad un OD di circa 0,5 - 0,6. Le cellule sono state raffreddate in ghiaccio per 30 minuti, poi centrifugate per 15 minuti a 5000 g a 4 °C. Una volta eliminato il sovrnatante, le cellule sono state risospese in una soluzione fredda contenente glicerolo sterile 10% (v/v) e, successivamente, centrifugate (15 minuti, 5000 g). La medesima operazione è stata effettuata diverse volte per eliminare i sali presenti nel terreno di coltura iniziale. Infine il pellet è stato risospeso in 2 ml della stessa soluzione e suddiviso in aliquote da 40 µl conservate a -80 °C.

## 2.7 Protocollo di trasformazione (elettroporazione)

Le aliquote di cellule competenti, scongelate in ghiaccio, sono state trasformate con circa 2 µl della miscela di ligazione. La sospensione di cellule e DNA è stata trasferita in apposite cuvette (0.1 cm) a cui è stato applicato un impulso di potenziale (1.8 kV, 200 ohm) mediante elettroporatore "Gene Pulser II" (Bio-Rad). Una volta verificata la correttezza della costante di tempo fornita dallo strumento è stato aggiunto rapidamente terreno SOC previamente raffreddato in ghiaccio; la sospensione così ottenuta è stata trasferita in una provetta Falcon da 15 ml. Dopo un'incubazione di circa un'ora a 37 °C in agitazione a 180 rpm, le cellule sono state trasferite su terreno solido addizionato con gli appropriati antibiotici. Come procedura abituale sono state ricavate due piastre da ogni campione sottoposto ad elettroporazione, una contenente 100 µl della coltura ed una con le restanti cellule, recuperate con 100 µl di sovrnatante dopo essere state centrifugate 2 minuti a 10000 g.

## 2.9 Screening delle colonie trasformate

Singole colonie sono state trasferite dalle piastre Petri in 50 µl di acqua sterile e disciolte tramite agitazione con il Vortex. Le cellule sono state sottoposte a bollitura in un forno a microonde (Moulinex, 900 Watts per 5')

e sedimentate tramite centrifugazione. Aliquote di 5  $\mu$ l dei supernatanti così ottenuti sono state sottoposte ad amplificazione coi seguenti primers, specifici per il vettore pBAD:

For: 5' GATTAGCGGATCCTACCTG 3'

Rev: 5' CACTACCATCGGCGCTACG 3'

La reazione di PCR è stata condotta utilizzando *Taq* DNA Polymerase (Recombinant, Fermentas), ed il seguente ciclo termico:

95 °C  $\times$  5 min;

95 °C  $\times$  45 sec; 58 °C  $\times$  30 sec; 72 °C  $\times$  60 sec \*(lunghezza dell'amplificato di 1 Kb) per 35 cicli

72 °C  $\times$  7 min

Infine i prodotti di PCR sono stati analizzati mediante elettroforesi (§2.3)

Il DNA dei cloni positivi è stato successivamente sottoposto a sequenziamento <sup>2</sup>.

## 2.10 Midiprep di DNA plasmidico

L'estrazione del DNA plasmidico è stata effettuata tramite il kit Perfect prep Plasmid Midi (Eppendorf). Il pellet di DNA purificato è stato risospeso in 120  $\mu$ l di acqua sterile. La quantificazione del DNA ottenuto è stata eseguita spettrofotometricamente.

## 2.11 Estrazione delle proteine totali dalle colture cellulari

Le cellule batteriche sono state recuperate per centrifugazione (10000 g per 30 min) e risospese in 5 ml di tampone di lisi (Tris-HCl 50 mM, NaCl 150 mM, PMSF 1 mM, pH 8). La lisi cellulare è stata ottenuta per sonicazione, in ghiaccio per tre minuti totali (impulsi di 30 sec intervallati da pause di 30 sec), con un sonicatore W-385 Ultrasonics utilizzato ad una potenza di 6 W. Il lisato è stato centrifugato a 10000 g per 20 minuti, e sottoposto a filtrazione (filtri con poro di 0,20  $\mu$ m, Albet Jacs).

## 2.12 Quantificazione delle proteine estratte

Le proteine totali contenute nei vari campioni esaminati sono state quantificate con il metodo descritto da Bradford (Bradford, 1976). Il metodo Bradford sfrutta l'effetto ipercromico a 595 nm del colorante Coomassie Blue G-250 in seguito alla formazione di complessi con le proteine. Per la retta di taratura sono state usate

---

<sup>2</sup> Il sequenziamento dei costrutti è stato effettuato nei laboratori BMR genomics (Spin-off ufficiale dell'Università di Padova)

diluizioni opportune di una soluzione di BSA (bovine-serum-albumin, Sigma) alla concentrazione di 1 mg/ml. Ogni determinazione è stata eseguita in doppio.

## **2.13 Gel filtrazione**

La cromatografia ad esclusione molecolare è stata effettuata con una colonna Superdex-200 (prep. grade, Amersham Pharmacia Biotech) di dimensioni 16 mm x 70 cm (volume totale 140 ml), provvista di un iniettore da 1ml, al flusso costante di 0,6 ml/ min. Il tampone di eluizione utilizzato è il seguente: Tris-HCl 50 mM, NaCl 150 mM, pH 8. Tutte le operazioni sono state svolte mantenendo costante la temperatura della colonna, pari a +4 °C e sono state raccolte frazioni di 0,9 ml, conservate a – 20 °C.

## **2.14 Analisi SDS-PAGE**

200 µl delle frazioni eluite dalla gel filtrazione (contenenti circa 30 - 100 µg di proteine) sono state precipitate in un uguale volume di una soluzione contenente il 20% di Acido Tricloroacetico (Sigma). Dopo 1 ora di incubazione in ghiaccio, i campioni sono stati centrifugati a 16000 g per 15 min. I pellet sono stati lavati in acetone e poi risospesi in tampone di caricamento. L'elettroforesi di proteine è stata condotta in gel di acrilamide/ bis-acrilamide 15%, Tris HCl pH 8.8 0.25 M, SDS 0.2 %, TEMED (0.1%), APS (0.25%) (stacking gel 4 %) in condizioni denaturanti (gel contenente 0,1 % di SDS) in celle Mini-Protean (Bio-Rad) per circa 2 ore a voltaggio costante (140 V) utilizzando un tampone di corsa contenente Tris 25 mM, glicina 200 mM e SDS 5 mM. I campioni sono stati sospesi in 20 µl di tampone per il caricamento in gel (Tris HCl 500 mM, pH 6.8, glicerolo 2.7 M, SDS 10%, blu di bromofenolo 0.5% e β-Mercaptoetanol 5%) quindi denaturati, prima del caricamento, mediante bollitura per 10 min. Le bande proteiche sono state evidenziate con Coomassie brilliant blu R-250.

## **2.15 Analisi Western blot**

Le proteine separate elettroforeticamente in SDS - PAGE sono state trasferite su di una membrana (Bio-rad) di PVDF (polyvinylidene difluoride). Il trasferimento su membrana di PVDF è stato eseguito elettroforeticamente mediante Mini-PROTEAN 3 Cell (Bio-Rad) in tampone di trasferimento (Tris 25 mM, glicina 192 mM) per 50 minuti a 80 V. Dopo l'adsorbimento dell'antigene, i restanti siti di legame per le



### 3.2 Clonaggio di forme tronche di *dnaQ*

Utilizzando 10 ng di pBAD- $\epsilon$ 243 come stampo, sono state amplificate mediante PCR alcune varianti tronche di *dnaQ* utilizzando primers specifici,

Primers	Sequenza 5'-3'
$\epsilon$ -186 rev	GATTCTGCAGTTAAGCCATCGACGTTTGACCACC
$\epsilon$ -213 rev	CGTGATCTGCAGTTAGCGTAACTTACTTGCCTGACG
$\epsilon$ -186, $\epsilon$ -213, $\epsilon$ -228, $\epsilon$ -234, $\epsilon$ -200 for	GGAATTAACCATGGGCACTGCAATTAC
$\epsilon$ -228 rev	CGGTCTGCAGGTATTTAACGGGCTTCATGAGCTGC
$\epsilon$ -234 rev	CGGACTGCAGGTATTTACTTCTGCACCAGATCGAGAC
$\epsilon$ -200 rev	CGTGCGCTGCAGTTATGCTTCACCTTGTTGCTGTTG

2.5 Unità dell' enzima Pfu Ultra High Fidelity DNA polymerase (Fermentas) e i seguenti cicli

95 °C  $\times$  2 min

95 °C  $\times$  30 sec; **Tm** °C  $\times$  40 sec; 72 °C  $\times$  **t** sec per 30 cicli

72 °C  $\times$  10 min

Dove

**Tm** è 65 °C per  $\epsilon$ -186,  $\epsilon$ -200,  $\epsilon$ -228,  $\epsilon$ -234; 66 °C per  $\epsilon$ -213

**t** è 40 sec per  $\epsilon$ -186, 41 sec per  $\epsilon$ -200, 42 sec per  $\epsilon$ -213, 45 sec per  $\epsilon$ -228,  $\epsilon$ -234,

I diversi frammenti amplificati sono stati digeriti con gli enzimi *Nco* I e *Pst* I (NEB) e clonati nel vettore di espressione pBAD ottenendo così i costrutti pBAD - $\epsilon$ 186, - $\epsilon$ 200, - $\epsilon$ 213, - $\epsilon$ 228, - $\epsilon$ 234, utilizzati per trasformare cellule elettrocompetenti di *E. coli* TOP10.

### 3.3 Coltura, lisi, estrazione delle proteine totali e gel filtrazione

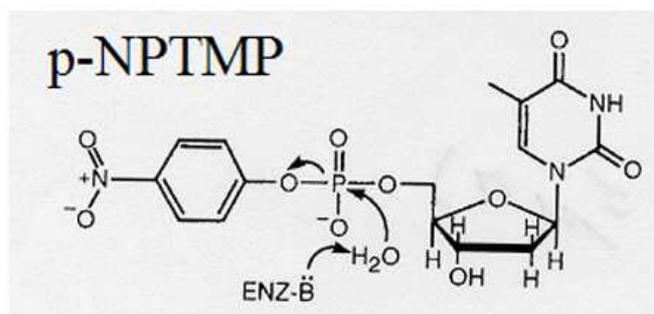
I ceppi di *Escherichia coli* TOP10/pBAD- $\epsilon$ 243, - $\epsilon$ 186, - $\epsilon$ 200, - $\epsilon$ 213, - $\epsilon$ 228, - $\epsilon$ 234 sono stati coltivati a 37 °C in preculture di 5 mL con terreno LB, addizionato di ampicillina. Tali preculture sono state diluite 1:500 in beute contenenti lo stesso terreno, in un volume totale di 100 ml. In corrispondenza di una densità di popolazione

equivalente ad un'assorbanza misurata a 600 nm pari a circa 0.6-0.7, il mezzo di coltura è stato addizionato di arabinosio 1 mM. Dopo 3 ore di induzione, le cellule sono state recuperate per centrifugazione. Quindi le cellule sono state lisate e le proteine totali sono state estratte (§ 2.11) e quantificate (§ 2.12)

4mg di proteine totali estratte sono state iniettate in una colonna cromatografica ad esclusione molecolare (§ 2.13).

### 3.4 Saggio di attività esonucleasica

Con le frazioni raccolte dalla colonna cromatografia ad esclusione molecolare sono state effettuate alcune determinazioni di attività esonucleasica. DnaQ, come altre proteine dotate di attività esonucleasica, catalizza l'idrolisi del 5'-p-nitrofenil estere della timidina 5'-monofosfato (pNP-TMP). Dalla reazione esonucleasica si forma il p-nitrofenolato che provoca un forte aumento dell'assorbanza a 420 nm, consentendo la determinazione dell'attività esonucleasica in presenza di enzima anche a concentrazioni minori di 100 nM (Hamdan, 2002).



Per i saggi di attività esonucleasica sono state utilizzate aliquote delle frazioni eluite dalla colonna di gel filtrazione. La soluzione per la determinazione dell'attività enzimatica è composta da:

140  $\mu\text{l}$  della frazione

20  $\mu\text{l}$  di tampone (NaCl 1,5 M, Tris-HCl 500 mM, DTT 10 mM, pH 8).

20  $\mu\text{l}$  di  $\text{MnCl}_2$  10 mM.

20  $\mu\text{l}$  di pNP-TMP 30 mM (Sigma)

Il substrato è stato aggiunto per ultimo dopo aver lasciato equilibrare la soluzione per 20 minuti. Le misure di assorbanza a 420 nm sono state effettuate ad intervalli di 2 minuti con un lettore di micropiastre Biorad 550 Microplate Reader.

### **3.5 Analisi SDS-PAGE e Western blot**

Con le frazioni eluite dalla gel filtrazione sono state condotte analisi SDS-PAGE e Western blot come riportato nei § 2.14 e 2.15.

## **4. Impiego di ceppi deleti per proteasi e chaperones molecolari**

### **4.1 Ceppi provenienti dalla Collezione Keio**

Alcuni ceppi di *E.coli* K-12 deleti per geni non essenziali codificanti per alcune proteasi o per alcuni chaperones molecolari sono stati ottenuti dall'*Escherichia coli* Genetic Stock Center di Yale.

I ceppi utilizzati sono indicati nel § 1.1. Prima del loro utilizzo per la sovraespressione del gene *dnaQ*, in tali ceppi è stato trasdotto l'allele *recA1907::cat-aadA* per evitare la ricombinazione omologa tra il gene *dnaQ* plasmidico e quello posizionato nel cromosoma.

### **4.2 Preparazione della stock fagica di P1 *vir* contenente l'allele *recA1907::cat-aadA***

Un lisato fagico di P1 *vir* c'è stato gentilmente fornito dal professor Gianni Dehò.

Una coltura overnight del ceppo BW11334/pBW120 (ceppo donatore) è stata diluita 1:100 in un volume finale di 2.5 ml di LB fresco addizionato di CaCl<sub>2</sub> 5 mM e glucosio 0.2 % e coltivata a 37 °C per 1 ora. La coltura batterica è stata infettata con 100 µl di sospensione fagica P1 *vir* (3x10<sup>9</sup> pfu ml<sup>-1</sup>) ed incubata a 37 °C per circa 3 ore. Trascorso questo tempo la coltura batterica è risultata completamente lisata dal fago. Sono state quindi aggiunte alcune gocce di cloroformio, il lisato è stato miscelato con il vortex e centrifugato a 14,000 rpm per 2 minuti. Il surnatante è stato conservato a 4 °C dopo l'aggiunta di 300 µl di cloroformio (Miller J. H., 1992).

### **4.3 Trasduzione dell'allele *recA1907::cat-aadA* mediata dal fago P1 *vir***

I ceppi di *E. coli* JW0429-1, JW0427-1, JW3903-5, JW0013-3, JW4103-1, JW0462-1, JW0426-1, provenienti dalla collezione Keio, sono stati coltivati overnight in 2 ml di terreno LB addizionato di kanamicina a 37 °C. Le

cellule sono state raccolte per centrifugazione (6000 g per 2') e risospese in un volume finale di LB fresco pari a quello iniziale al quale sono stati addizionati  $\text{MgSO}_4$  100 mM,  $\text{CaCl}_2$  5 mM e kanamicina. A 100  $\mu\text{l}$  di tali colture batteriche sono stati aggiunti 100  $\mu\text{l}$  del lisato fagico contenente l'allele *recA1907::cat-aadA* e incubati a 37 °C per 30'. Sono stati aggiunti 200  $\mu\text{l}$  di Na-citrato 1 M (pH 5.5) e 1 ml di LB fresco e la soluzione contenente fagi e batteri è stata incubata a 37 °C per 1 ora. Le cellule sono state raccolte per centrifugazione, risospese in 100  $\mu\text{l}$  di LB addizionato di Na-citrato 100 mM (pH 5.5) e piastrate su piastre contenenti gli antibiotici selettivi cloramfenicolo e kanamicina.

#### **4.4 Trasformazione con il costrutto pBAD- $\epsilon$ 243**

Singole colonie di JW0429-1, JW0427-1, JW3903-5, JW0013-3, JW4103-1, JW0462-1, JW0426-1 trasdotte con l'allele *recA1907::cat-aadA* sono state purificate tramite streaking e propagate in piastre di LB addizionato con kanamicina e cloramfenicolo.

Sono state quindi preparate cellule elettrocompetenti (§ 2.6) di tali ceppi batterici trasdotti, che sono state in seguito elettroporate con il costrutto pBAD- $\epsilon$ 243.

#### **4.5 Coltura, lisi, estrazione di proteine, gel filtrazione e analisi Western blot**

I ceppi *Escherichia coli* JW0429-1, JW0427-1, JW3903-5, JW0013-3, JW4103-1, JW0462-1, JW0426-1 trasdotti con l'allele *recA1907::cat-aadA* sono stati coltivati a 37 °C in preculture di 5 mL con terreno LB, addizionate di kanamicina e cloramfenicolo. Tali preculture sono state diluite 1:500 in beute contenenti lo stesso terreno, in un volume totale di 100 mL. In corrispondenza di una densità di popolazione equivalente ad un'assorbanza a 600 nm pari a circa 0.6- 0.7 il mezzo di coltura è stato addizionato di arabinosio 1 mM. Dopo 3 ore di induzione, le cellule sono state recuperate per centrifugazione. Quindi le cellule sono state lisate, le proteine totali sono state estratte (§ 2.11) e quantificate (§ 2.12)

4mg di proteine totali estratte sono state iniettate in una colonna cromatografica ad esclusione molecolare (§ 2.13).

Con le frazioni eluite dalla gel filtrazione sono state condotte analisi Western blot come riportato nel § 2.15.

### **5. Realizzazione e impiego del costrutto pBAD-*dnaE***

#### **5.1 Clonaggio di *dnaE***

300 ng di DNA genomico estratti da cellule di *E.coli* TOP10 sono stati utilizzati per amplificare, attraverso PCR, il gene *dnaE* di 3800 bp utilizzando i primers

*dnaE* for GGGTACCCAACCGCTTTTCATCTGGAAA

*dnaE* rev CCATGCACCGAGATCGGCGAAGATTTTAC,

l'enzima Pfu Ultra II Fusion HS DNA polymerase (Stratagene) e il seguente ciclo

95 °C × 5 min;

95 °C × 30 sec; 60 °C × 30 sec; 72 °C × 1min 30 sec per 30 cicli

72 °C × 7 min

Il prodotto di PCR è stato analizzato mediante elettroforesi in gel di agarosio, recuperato, purificato (§ 2.3 e 2.4) e clonato utilizzando il kit StrataClone Blunt PCR Cloning (Stratagene) seguendo le indicazioni fornite dalla ditta.

Il gene è stato sequenziato e in seguito amplificato utilizzando i primers

$\alpha$ -1160 rev 5' CGTGCTGCAGCAGACTCATAGTATTCCTG 3'

$\alpha$ -1160 for 5' CGATCCATGGGTGAACCACGTTTCGTACAC 3'

l'enzima Pfu Ultra II Fusion HS DNA polymerase (Stratagene) e il seguente ciclo

95 °C × 5 min;

95 °C × 30 sec; 60 °C × 30 sec; 72 °C × 1 min 23 sec per 30 cicli

72 °C × 7 min

Infine il prodotto di PCR è stato analizzato mediante elettroforesi in gel di agarosio, recuperato e purificato (§ 2.3 e 2.4). Tale frammento e il vettore di espressione pBAD sono stati digeriti con gli enzimi *Nco* I e *Pst* I (NEB), quindi è stata effettuata una reazione di ligazione (§ 2.5) ottenendo così il costrutto pBAD- $\alpha$ 1160, utilizzato per trasformare cellule elettrocompetenti di *E.coli* TOP10.

## 5.2 Clonaggio forme tronche di *dnaE*

Utilizzando 10 ng di pBAD- $\alpha$ 1160 come stampo, alcune varianti tronche di *dnaE* sono state amplificate attraverso PCR utilizzando primers specifici qui di seguito elencati.

Primers	Sequenza 5'-3'
$\alpha$ -287, $\alpha$ -308, $\alpha$ -336, $\alpha$ -353 for	GAATTAACCATGGGTGAACCACGTTTCGTACAC
$\alpha$ -287 rev	GAATCTGCAGTTAGCTCATGTCCCCGGTCGGGAACTGC
$\alpha$ -308 rev	CGCGATCTGCAGTTAGAATAAAAAGGCCAGACGCTC
$\alpha$ -336 rev	CGTGATCTGCAGTTAGCCCATCTGGTTGATAACCTG
$\alpha$ -353 rev	CGGTCTGCAGTCAATCTTTCGACCACTGGATAAATTC

Inoltre, per la reazione PCR sono state utilizzate 2.5 unità dell'enzima Pfu Ultra II Fusion HS DNA polymerase (Stratagene) e i seguenti cicli di amplificazione:

95 °C × 5 min

95 °C × 30 sec; 65 °C × 30 sec; 72 °C × **t** sec per 30 cicli

72 °C × 7 min

Dove

**t** è 14 sec per  $\alpha$ -287, 15 sec per  $\alpha$ -308, 16 sec per  $\alpha$ -336 e 20 sec  $\alpha$ -353

I diversi frammenti amplificati sono stati digeriti con gli enzimi *NcoI* e *PstI* (NEB) e clonati nel vettore di espressione pBAD ottenendo così i costrutti pBAD- $\alpha$ 287, - $\alpha$ 308, - $\alpha$ 336, - $\alpha$ 353.

Tutti questi costrutti, il vettore pBAD vuoto e pBAD- $\alpha$ 1160 sono stati utilizzati per trasformare il ceppo *E. coli* BW25993, nel quale era stato introdotto l'allele *recA1907::cat-aadA* attraverso trasduzione mediata dal fago P1 (§4.3).

Il fenotipo di questi costrutti è stato analizzato per 30-50 ore attraverso lo streaking di singole colonie su piastre Petri contenenti LB-agar addizionato di ampicillina, isopropil- $\beta$ -tiogalattopiranoside (Shelton) 1 mM, arabinosio 13 mM e 5-bromo-4-cloro-3-indolil- $\beta$ -D-galattopiranoside (X-Gal) 40  $\mu$ g ml<sup>-1</sup>.

## 6. Realizzazione e impiego del costrutto pBAD-His-dnaQ

### 6.1 Clonaggio di His- *dnaQ*

Utilizzando 10 ng di pBAD- $\epsilon$ 243 come stampo, è stato amplificato mediante PCR il gene *dnaQ* per il clonaggio nel vettore pBAD-His utilizzando i seguenti primers,

pBAD-His $\epsilon$ 243 for        5' CGTACTCGAGAATGAGCACTGCAATTACACGCC        3'

pBAD-His $\epsilon$ 243 rev        5' GTATCTGCAGGTATTTATGCTCGCCAGAGGCAAC        3'

2.5 unità dell'enzima Pfu Turbo DNA polymerase (Stratagene) e il seguente ciclo

95 °C  $\times$  3 min;

95 °C  $\times$  32 sec; 69 °C  $\times$  40 sec; 72 °C  $\times$  50 sec    per 32 cicli

72 °C  $\times$  10 min

Il prodotto di PCR è stato analizzato mediante elettroforesi in gel di agarosio, recuperato e purificato (§ 2.3 e 2.4)

Il frammento amplificato e il vettore di espressione pBAD sono stati digeriti con gli enzimi *Xho* I e *Pst* I (NEB), quindi è stata effettuata una reazione di ligazione (§ 2.5) ottenendo così il costrutto pBAD-His- $\epsilon$ 243 che è stato sequenziato ed utilizzato per trasformare cellule elettrocompetenti di *E. coli* TOP10.

### 6.2 Coltura e lisi

Il ceppo *Escherichia coli* TOP10/pBAD-His- $\epsilon$ 243 è stato coltivato a 37 °C in precolture di 5 ml con terreno LB, addizionato di ampicillina. Tali precolture sono state diluite 1:100 in beute contenenti lo stesso terreno, in un volume totale di 800 ml. In corrispondenza di una densità di popolazione equivalente ad un'assorbanza a 600 nm pari a circa 0.6- 0.7 il mezzo di coltura è stato addizionato di arabinosio 1 mM. Dopo 3 ore di induzione, le cellule sono state recuperate per centrifugazione e risospese in 40 ml di tampone di lisi (Tris-HCl 50 mM, NaCl 0,5 M, PMSF 1 mM, pH 8) e lisate per sonicazione, in ghiaccio per tre minuti totali (impulsi di 30 secondi intervallati da pause di 30 secondi).

### 6.3 Cromatografia d'affinità

La colonna d'affinità "HiTrap Chelating Affinity" (Amersham Pharmacia Biotech, dimensioni 1,6 x 2,5 cm, volume 5 ml), interfacciata ad un cromatografo Gradi-Frac (Amersham Pharmacia Biotech), è stata utilizzata per separare DnaQ contenente un motivo esaistidinico all'N-terminale dalle restanti proteine estratte. Il flusso

d'esercizio utilizzato, dove non specificato diversamente, è stato imposto ad un regime pari a 1 ml/minuto. La colonna è stata lavata con 25 ml d'acqua; successivamente quest'ultima è stata equilibrata con il medesimo volume di una soluzione di NiSO<sub>4</sub> (0,1 M). La colonna è stata poi condizionata per 30- 40 minuti con tampone Tris 50 mM, NaCl 500 mM, pH 8, prima del caricamento delle proteine. Le proteine estratte dal campione (40 ml) sono state caricate in colonna (flusso 0,5 ml/min) e la concentrazione di proteina in uscita monitorata spettrofotometricamente a 280 nm.

La colonna caricata è stata lavata con tampone fino alla scomparsa di proteina in uscita. Durante il caricamento sono state raccolte frazioni da 1,5 ml. L'eluizione è stata effettuata con un gradiente di imidazolo (da 0 a 0.5 M, 10 volumi di colonna) e l'eluato è stato raccolto in frazioni da 0.8 ml. La proteina d'interesse è stata eluita ad una concentrazione d'imidazolo pari a circa 250 mM. Sulle frazioni raccolte sono stati effettuati saggi di attività esonucleasica (§ 3.5) e le migliori sono state concentrate con una cella per ultrafiltrazione Amicon fino a ridurre il volume a 2 ml.

## **6.4 Spettrometria di massa**

Le bande di interesse evidenziate mediante SDS-page sono state recuperate dal gel di poliacrilammide ed inviate al laboratorio di Proteomica dell'Università di Salerno per l'identificazione mediante spettrometria di massa.

Per l'identificazione finale della proteina a partire dai frammenti peptidici generati dall'analisi è stato utilizzato il seguente programma:

<http://www.matrixscience.com/>

## **7. Utilizzazione dell'estratto proteico di cellule TOP10**

### **7.1 Coltura, lisi e gel filtrazione**

Il ceppo *Escherichia coli* TOP10 è stato coltivato a 37 °C in precoltura di 5 ml con terreno LB. Tale precoltura è stata diluita 1:100 in beute contenenti lo stesso terreno, in un volume totale di 100 ml. In corrispondenza di una densità di popolazione equivalente ad un'assorbanza a 600 nm pari a circa 0.6- 0.7 le cellule sono state recuperate per centrifugazione.

Le proteine totali sono state estratte (§ 2.10) e 4mg di campione sono stati iniettati in una colonna cromatografica ad esclusione molecolare (§ 1.13).

## 7.2 Gel bidimensionale e analisi attraverso spettrometria di massa

200  $\mu$ l (concentrazione di circa 30 - 100  $\mu$ g ) delle frazioni eluite dalla gel filtrazione sono state precipitate in un uguale volume di una soluzione contenente il 20% di Acido Tricloroacetico (Sigma). Dopo 1 h di incubazione in ghiaccio, i campioni sono stati centrifugati a 16000 g per 15 minuti. I pellets sono stati lavati in acetone e poi risospesi in 200  $\mu$ l di buffer di idratazione (urea 7 M, tiourea 2M, CHAPS 2%, DTT 9 mg/ml) per 12 ore. Il campione è stato centrifugato per 15 minuti a 8 °C a 16000g.

150  $\mu$ l di campione ai quali sono stati aggiunti 2.3  $\mu$ l di DTT (400 mg/ml), 3  $\mu$ l di IPG buffer pH 4-7 (Amersham Pharmacia Biotech) e 1  $\mu$ l di Blu di Bromofenolo sono stati utilizzati per idratare over-night una striscia di poliacrilamide da 7 cm a pH 4- 7 (Amersham Pharmacia Biotech).

La isofocalizzazione è stata effettuata in Protean IEF Cell (Bio Rad) alle seguenti condizioni:

300 V per 1 ora

1000 V per 4 ore

3000 V per 4 ore

500 V fino allo spegnimento dello strumento

La striscia contenete le proteine isofocalizzate è stata equilibrata in tampone Tris acetato 50 mM, SDS 2%, Urea 6 M, glicerolo 30%, pH 6.8 per 30 minuti, in seguito è stata aggiunta iodoacetamide (1.5%) per sostituire il DTT. La striscia è stata poi caricata su un gel di Acrilamide/ bis-acrilamide 10%, Tris HCl pH 8.8 0.25 M, SDS 0.1 %, TEMED (0.05 %), APS (0.25%).

La corsa della seconda dimensione è stata condotta utilizzando celle Mini-Protean (Bio-Rad). L'elettroforesi è stata condotta per circa 2 ore a voltaggio costante (50 V) utilizzando il tampone 1 $\times$  Tris 25 mM, acido acetico 15 mM per l'interno della cella elettroforetica e un tampone 1 $\times$  Tris 200 mM, SDS 14 mM, Tricina 120 Mm per la parte esterna.

Il gel è stato colorato con Coomassie brilliant blu R-250 in 40%v/v metanolo-10% v/v acido acetico, in agitazione per circa un'ora.

Gli spots di interesse sono stati escissi dal gel ed analizzati. (§ 6.6).

## 8. Utilizzazione del ceppo di *Escherichia coli* TOP10/pBAD-AR1

Il ceppo di *E. coli* TOP 10/pBAD-AR1 disponibile nel nostro laboratorio, (tesi di dottorato di Alessandra Stefan "Analisi strutturale dei determinanti del silencing indotto da RNA antisenso in *Escherichia coli* ". Università

degli studi di Bologna, XVII ciclo) se opportunamente indotto con arabinosio, codifica per un RNA antisenso in grado di appaiarsi alla regione regolativa di 221 nucleotidi dell'RNA prodotto dalla trascrizione del gene *dnaQ*.

### **8.1 Coltura, lisi, estrazione delle proteine totali**

Il ceppo di *Escherichia coli* TOP10/pBAD-AR1 è stato coltivato a 37 °C in precolture di 5 ml con terreno LB, addizionato di ampicillina. Tali precolture sono state diluite 1:1000 in beute contenenti lo stesso terreno, in un volume totale di 800 ml.

Dopo 6 ore le cellule sono state recuperate per centrifugazione oppure indotte per 15 ore con arabinosio 13 mM e quindi raccolte. Nei due casi, le cellule sono state lisate e le proteine totali sono state estratte (§ 2.11) e quantificate (§ 2.12).

### **8.2 Gel filtrazione e saggio di attività esonucleasica**

La cromatografia è stata effettuata utilizzando le proteine solubili totali estratte da TOP10/pBAD-AR1 inducendo o meno l'antisense. In entrambi i casi, 4mg di proteine estratte sono state iniettate in una colonna cromatografica ad esclusione molecolare (§ 2.13).

Con le frazioni raccolte dalla colonna cromatografia ad esclusione molecolare sono state effettuate alcune determinazioni di attività esonucleasica (§ 3.4).

# RISULTATI E DISCUSSIONE

## 1. Impiego del costrutto pBAD-dnaQ e delle forme tronche di tale gene

### 1.1 Clonaggio di *dnaQ* e di forme tronche del gene

La subunità  $\epsilon$  della Pol III è intrinsecamente instabile (Foster e Marinus, 1992) e soggetta a proteolisi. La regione C-terminale (aminoacidi 187-243) conferisce instabilità alla proteina (Perrino et al., 1999). Pertanto è stato realizzato un costrutto che contiene solo la regione N-terminale della proteina ovvero i primi 186 aminoacidi ( $\epsilon$ 186).

Inoltre sono state progettate delle forme tronche della subunità  $\epsilon$ , delete di una porzione più o meno ampia della regione C-terminale per analizzarne la stabilità e la capacità di associazione all'oloenzima Pol III.

Utilizzando l'algoritmo Psipred (Jones, 1999) è stato possibile costruire un modello della struttura secondaria della regione C-terminale; in particolare, in tale modello sono presenti due  $\alpha$  eliche nei residui compresi tra le posizioni 201-205 e i siti 219-234 (figura 1).

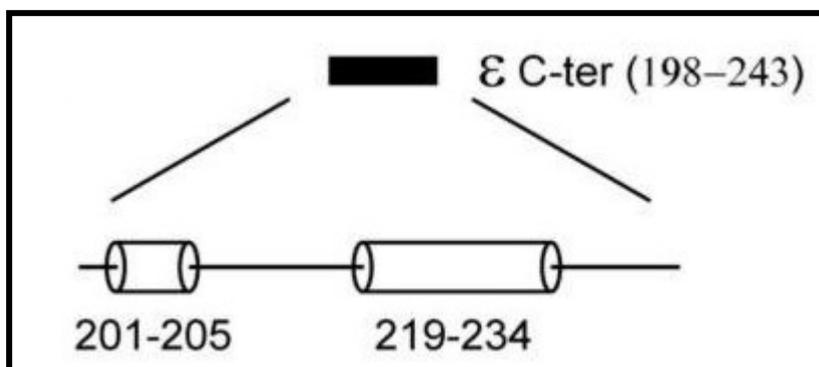


Fig 1: Struttura secondaria della regione C-terminale di  $\epsilon$  (residui 198-243). Sono presenti due  $\alpha$  eliche nei residui compresi tra le posizioni 201-205 e i siti 219-234.

Sono state realizzate delle varianti della subunità  $\epsilon$  con un diverso contenuto in  $\alpha$  eliche. In particolare i costrutti che sono stati utilizzati sono i seguenti (figura 2):

- pBAD- $\epsilon$ 213 (fornito di una sola  $\alpha$  elica)
- pBAD- $\epsilon$ 228 (fornito di una  $\alpha$  elica intera e con la seconda elica troncata)
- pBAD- $\epsilon$ 234 (fornito di entrambe le  $\alpha$  eliche)

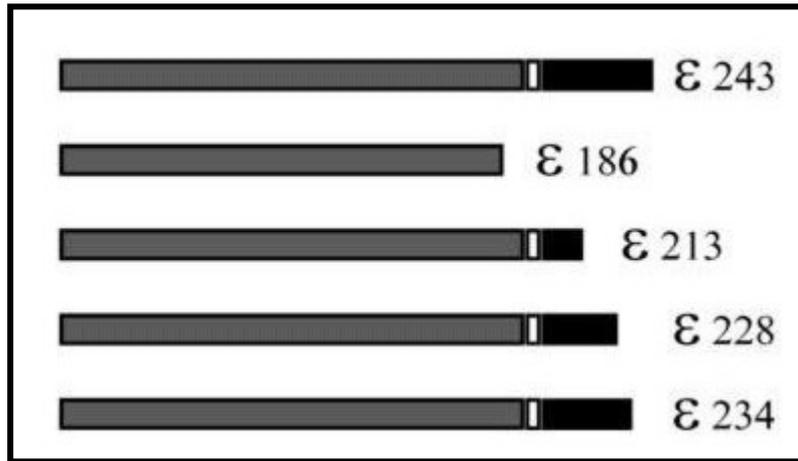


FIG 2:  $\epsilon$  intera ( $\epsilon$ 243) e alcune varianti che presentano una delezione più o meno ampia della regione C-terminale. Regione N-terminale catalitica (grigio), Q-linker (bianco), regione C-terminale (nero).

## 1.2 Gel filtrazione e saggio di attività esonucleasica

Dopo la sovraespressione delle varianti di  $\epsilon$ , le proteine cellulari sono state estratte e separate mediante gel filtrazione (cromatografia per esclusione molecolare). Per le proteine globulari, il tempo di ritenzione in colonna cromatografica dipende linearmente dalle dimensioni e quindi dalla massa molecolare delle proteine stesse. Una volta calibrata la colonna con proteine globulari di massa nota, è stato possibile dedurre la massa ignota di una proteina globulare sulla base del suo tempo di ritenzione in colonna. Le proteine con massa maggiore passano velocemente e compaiono nelle prime frazioni raccolte.

Per caratterizzare ogni variante di  $\epsilon$  sovraespressa, sono stati effettuati saggi di attività esonucleasica sulle 70 frazioni eluite dalla gel filtrazione, utilizzando il 5'-p-nitrofenil estere della timidina 5'-monofosfato (pNP-TMP). Sono stati evidenziati due picchi di attività esonucleasica nelle frazioni corrispondenti a proteine aventi massa molecolare di circa 30 kDa (frazione  $52 \pm 1$ ) (figura 3) e di circa 1000 kDa (frazione  $6 \pm 1$ ) (figura 4).

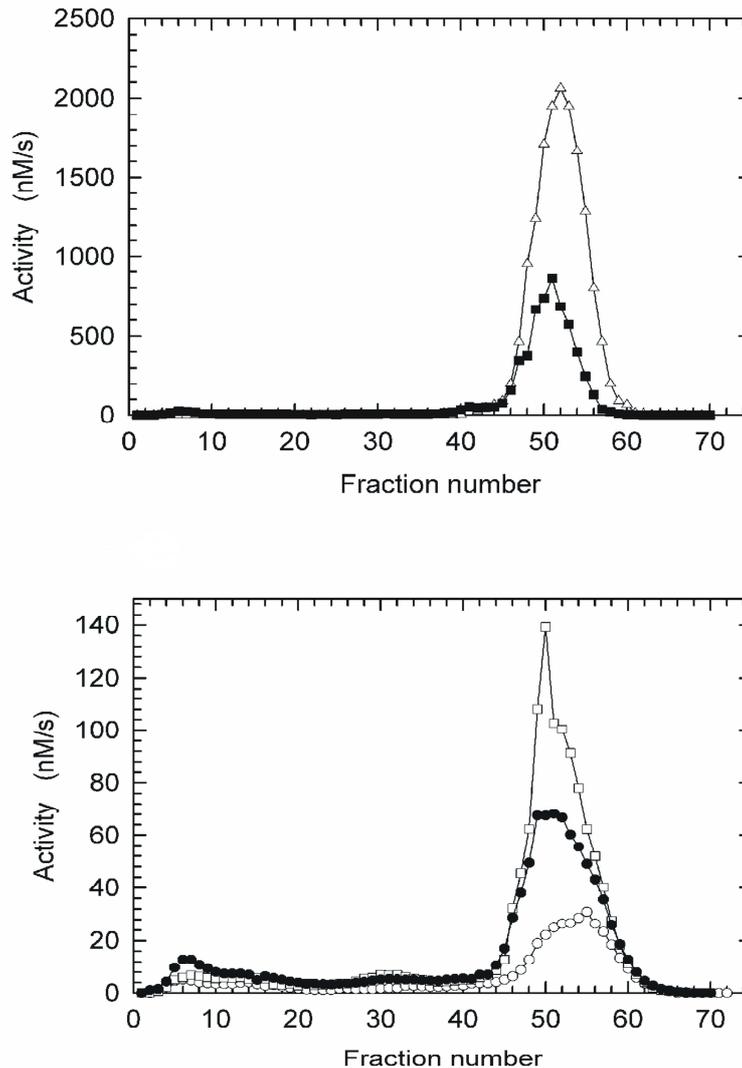


FIG 3: Attività esonucleasica delle frazioni eluite dalla GF dell'estratto proteico di cellule di TOP 10 nelle quali è stato sovraespresso il costrutto pBAD- $\epsilon$ 186 ( $\Delta$ ), pBAD- $\epsilon$ 213 ( $\blacksquare$ ), pBAD- $\epsilon$ 228 ( $\square$ ), pBAD- $\epsilon$ 234 ( $\bullet$ ), pBAD- $\epsilon$ 243 ( $\circ$ ). In entrambi i grafici si osserva un picco di attività centrato sulla frazione  $52 \pm 1$ .

In figura 3 è possibile osservare un picco di attività esonucleasica in corrispondenza della frazione corrispondente a proteine di massa molecolare compatibili con la forma monomerica di  $\epsilon$  (30 kDa). Nel caso di sovraespressione di  $\epsilon$ 186, il picco di attività è centrato in corrispondenza della frazione 52 e l'attività è maggiore rispetto ai valori osservati con le altre varianti di  $\epsilon$ . Un picco più basso di circa 2.4 volte, con un massimo nella frazione 51, è osservabile nel caso di sovraespressione di  $\epsilon$ 213. Nel caso di  $\epsilon$ 228,  $\epsilon$ 234,  $\epsilon$ 243 l'attività risulta essere rispettivamente circa 15, 30, 70 volte inferiore rispetto ad  $\epsilon$ 186. Il picco di attività risulta essere non simmetrico e allargato, elemento che suggerisce la presenza di forme diversamente degradate della variante di  $\epsilon$  sovraespressa. I dati relativi alla attività esonucleasica delle varianti troncate di  $\epsilon$  indicano che  $\epsilon$ 186 è, tra quelle esaminate, la più attiva. Questo potrebbe essere determinato dal fatto che  $\epsilon$ , privata

della sua regione C-terminale è molto stabile e quindi, essendo più abbondante rispetto alle altre varianti sovraespresse, produce il picco con la maggiore attività. Per lo stesso motivo si può ipotizzare che tanto più ampia è la regione C-terminale tanto più la proteina è instabile; questo spiegherebbe il progressivo calo di attività che si osserva all'aumentare della lunghezza del dominio non catalitico di  $\epsilon$ .

Nell'ambito delle misure di attività esonucleasica delle frazioni eluite dalla gel filtrazione è stato possibile registrare un picco di attività centrato in corrispondenza della frazione 6. Sono eluite in tale frazione le proteine con una massa molecolare di circa 1.0 MDa, che è compatibile con la presenza della forma di  $\epsilon$  complessata all'oloenzima Pol III. In figura 4 è rappresentato l'andamento dell'attività esonucleasica delle frazioni dalla 1 alla 30. La scala è diversa da quella della figura 3 per potere apprezzare le diverse altezze dei picchi di attività in corrispondenza della frazione 6.

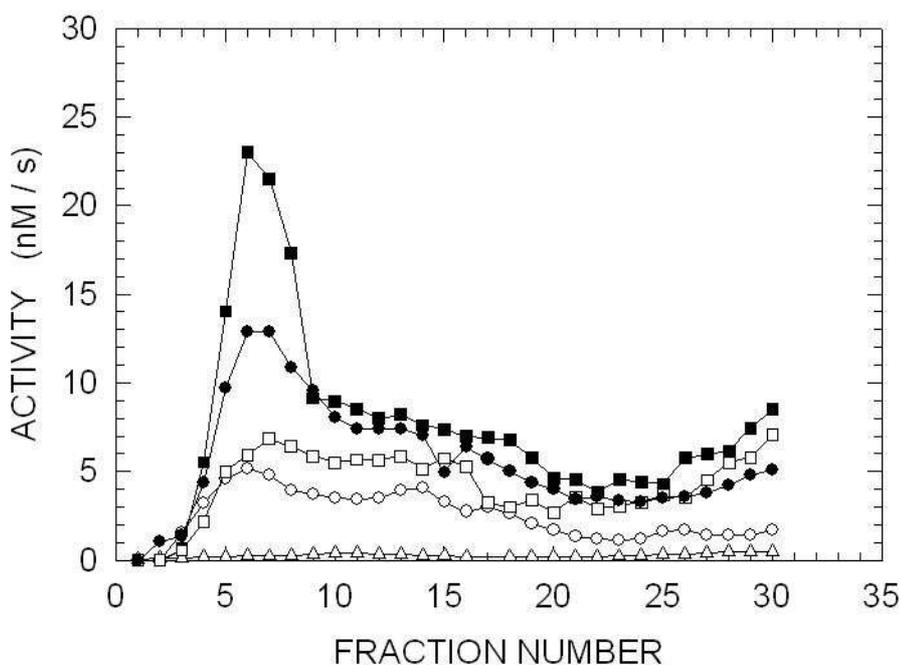


FIG 4 : Particolare dell' attività esonucleasica delle frazioni 1-30 eluite dalla GF. Simboli  $\Delta$  pBAD- $\epsilon$ 186,  $\blacksquare$  pBAD- $\epsilon$ 213,  $\square$  pBAD- $\epsilon$ 228,  $\bullet$  pBAD- $\epsilon$ 234,  $\circ$  pBAD- $\epsilon$ 243. Si può osservare un picco di attività centrato sulla frazione  $6 \pm 1$ .

In particolare, si osserva che l' attività è praticamente nulla nel caso di  $\epsilon$ 186 ed aumenta in maniera rilevante nel caso delle altre varianti di  $\epsilon$ , soprattutto nel caso di  $\epsilon$ 213, per la quale si osserva il picco di attività più intenso. L'attività esonucleasica delle altre varianti di  $\epsilon$  nelle frazioni dell'oloenzima assume valori intermedi. In particolare,  $\epsilon$ 243 risulta essere meno attiva rispetto a  $\epsilon$ 228 e  $\epsilon$ 234 e quest'ultima più attiva di  $\epsilon$ 228.

La attività quasi nulla di  $\epsilon 186$  suggerisce che tale variante di  $\epsilon$  non sia in grado di associarsi all'oloenzima poichè la perdita della regione C-terminale di  $\epsilon$  impedisce il legame con  $\alpha$  e pertanto l'assemblaggio di  $\epsilon 186$  nella Pol III. Quando viene sovraespressa, la variante  $\epsilon 186$  non viene quindi sequestrata nell'oloenzima e questo potrebbe fornire un'ulteriore spiegazione all' ampiezza del picco corrispondente alla forma monomericamente di  $\epsilon 186$  nel quale confluisce tutta la proteina sovraespressa.

Come si osserva in figura 4, il picco corrispondente alla variante  $\epsilon 213$  associata all'oloenzima risulta essere circa 20 volte più intenso di quello della variante  $\epsilon 186$ . Questo dato indica che  $\epsilon 213$ , contrariamente a quanto accade per  $\epsilon 186$ , è in grado di legarsi alla subunità polimerasica e di partecipare all' assemblaggio della Pol III. Evidentemente la regione C-terminale di  $\epsilon 213$  è sufficiente per il legame con  $\alpha$ .

Per quanto riguarda le altre varianti tronche e la loro associazione nell'oloenzima, il fatto che si possa registrare un' attività esonucleasica nelle rispettive frazioni 6 indica senza dubbio che, sovraespresprimendole, si vanno a complessare con la Pol III. La variante  $\epsilon 243$  associata nell'oloenzima risulta essere poco attiva (solo cinque volte più attiva di  $\epsilon 186$ ) e questo potrebbe dipendere dalla sua grande instabilità. Nelle misure di attività delle frazioni corrispondenti all'oloenzima, il picco di  $\epsilon 234$  risulta essere più intenso rispetto a quello di  $\epsilon 228$  probabilmente perché, essendo presente integralmente in  $\epsilon 234$  la  $\alpha$  elica II della regione C-terminale di  $\epsilon$  (residui 219-234), la proteina acquista stabilità. La variante  $\epsilon 228$  è invece troncata a livello di questa seconda  $\alpha$  elica e questo potrebbe determinare un indebolimento ed una destrutturazione della proteina, che perde parte della sua struttura terziaria.

In figura 5 sono riassunti gli andamenti di attività esaminati per quanto riguarda la subunità  $\epsilon$  monomericamente e la sua associazione nell'oloenzima. L'associazione di  $\epsilon$  con la Pol III potrebbe dipendere principalmente dai residui compresi tra gli aminoacidi 187 e 213. Ad eccezione di  $\epsilon 186$ , le altre varianti di  $\epsilon$  sovraesprese si associano alla subunità polimerasica e quindi partecipano all'assemblaggio dell'oloenzima. Tale partecipazione potrebbe dipendere principalmente dalla stabilità di queste varianti che è inversamente proporzionale alla lunghezza della regione C-terminale ( $\epsilon 243$  vs  $\epsilon 213$ ) oppure può dipendere dalla perdita della struttura tridimensionale della proteina ( $\epsilon 234$  vs  $\epsilon 228$ ).

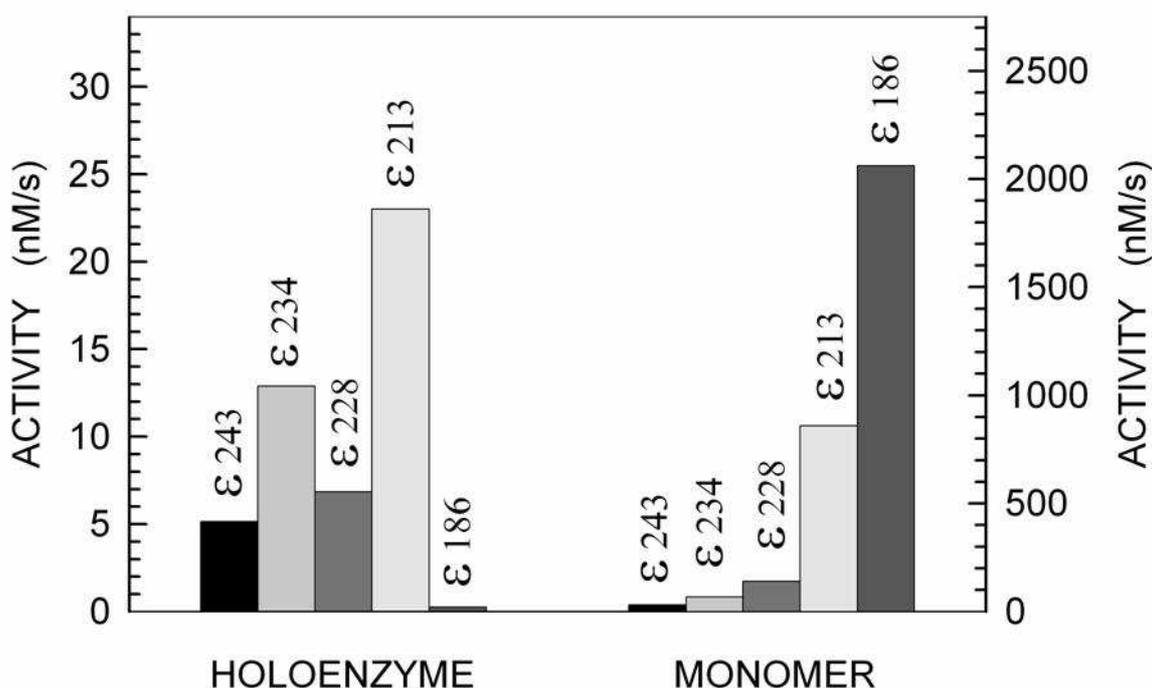


FIG 5: Diagramma riassuntivo di confronto dell' attività esonucleasica (nM/s) delle frazioni corrispondenti alla forma monomerica e alla forma associata all'oloenzima delle varianti di  $\epsilon$  sovraespresse.

Per ottenere ulteriori informazioni riguardo il coinvolgimento della regione compresa tra i residui 187-213 di  $\epsilon$  per il legame con  $\alpha$ , è stata progettata un' ulteriore variante di  $\epsilon$  priva di entrambe le  $\alpha$  eliche della regione C-terminale, ossia terminante con il residuo 200. In figura 6 è rappresentato l'andamento dell'attività esonucleasica delle frazioni eluite dalla gel filtrazione dell'estratto proteico di cellule TOP10/pBAD- $\epsilon$ 200. L'attività che si misura nella frazione corrispondente alla forma monomerica di  $\epsilon$ 200 ha un valore inferiore a quello di  $\epsilon$ 186 e di  $\epsilon$ 213, ma maggiore di quello riscontrato per la variante  $\epsilon$ 228.

Per quanto riguarda l'attività esonucleasica che si registra nella frazione corrispondente all' oloenzima quando  $\epsilon$ 200 è sovraespresa, essa risulta maggiore di quella misurata per  $\epsilon$ 186, ma comunque minore di quella determinata per  $\epsilon$ 213. Si può allora ipotizzare che i determinanti per il legame alla subunità polimerasica siano compresi tra i residui 201 e 213 di  $\epsilon$ .

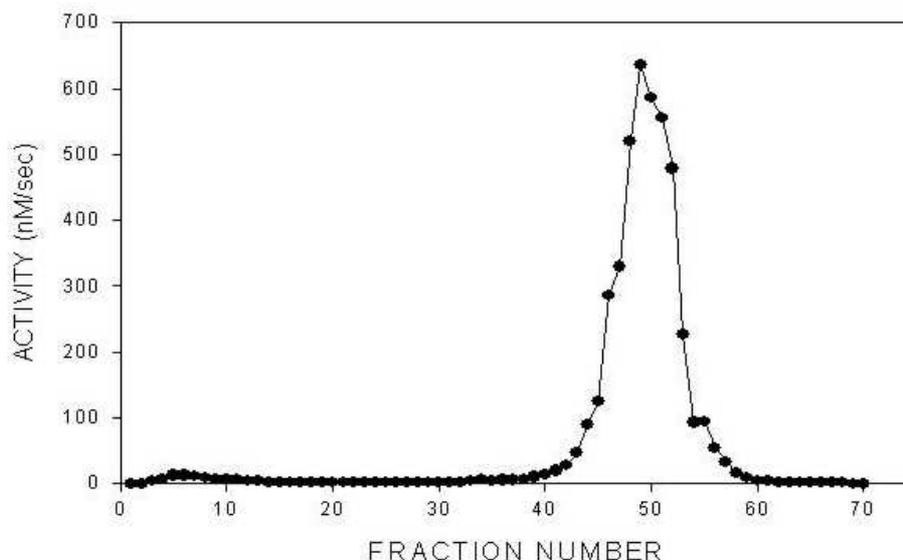


FIG 6 Attività esonucleasica delle frazioni eluite dalla GF dell'estratto proteico di cellule di TOP 10 nelle quali è stato sovraespresso il costrutto pBAD- $\epsilon$ 200.

## 1.4 Analisi SDS-PAGE e Western blotting

In figura 7 sono riportati i Western blotting, realizzati utilizzando un anticorpo anti  $\epsilon$ , di alcune frazioni isolate, come descritto precedentemente, mediante gel filtrazione.

La presenza di  $\epsilon$  è confermata nelle frazioni corrispondenti all'oloenzima (frazione 6) di tutte le varianti sovraesprese di  $\epsilon$ . Dal confronto tra le corsie corrispondenti alla frazione 6 di tutte le varianti di  $\epsilon$ , emerge chiaramente che la forma  $\epsilon$ 186 presenta una banda meno intensa, dato che confermerebbe i risultati di attività esonucleasica per i quali la attività della frazione dell'oloenzima, nel caso di sovraespressione di  $\epsilon$ 186, è praticamente nulla. Un' ulteriore conferma di questo dato proviene dall'analisi SDS-PAGE (figura 8) delle frazioni. La colorazione delle bande proteiche con Coomassie permette di evidenziare la presenza della proteina sovraespressa in tutte le corsie ad eccezione della prima (relativa alla frazione 6 dell'estratto proteico di  $\epsilon$ 186) e di quella relativa al controllo negativo. Pertanto è possibile affermare definitivamente che  $\epsilon$ 186 non è in grado di associarsi all'oloenzima perché mancante della regione C-terminale responsabile del legame alla subunità polimerasica.

Come appare in figura 7 e in figura 8, diversamente da  $\epsilon$ 186, le altre varianti di  $\epsilon$  sono presenti nella frazione 6 delle rispettive gel filtrazioni.

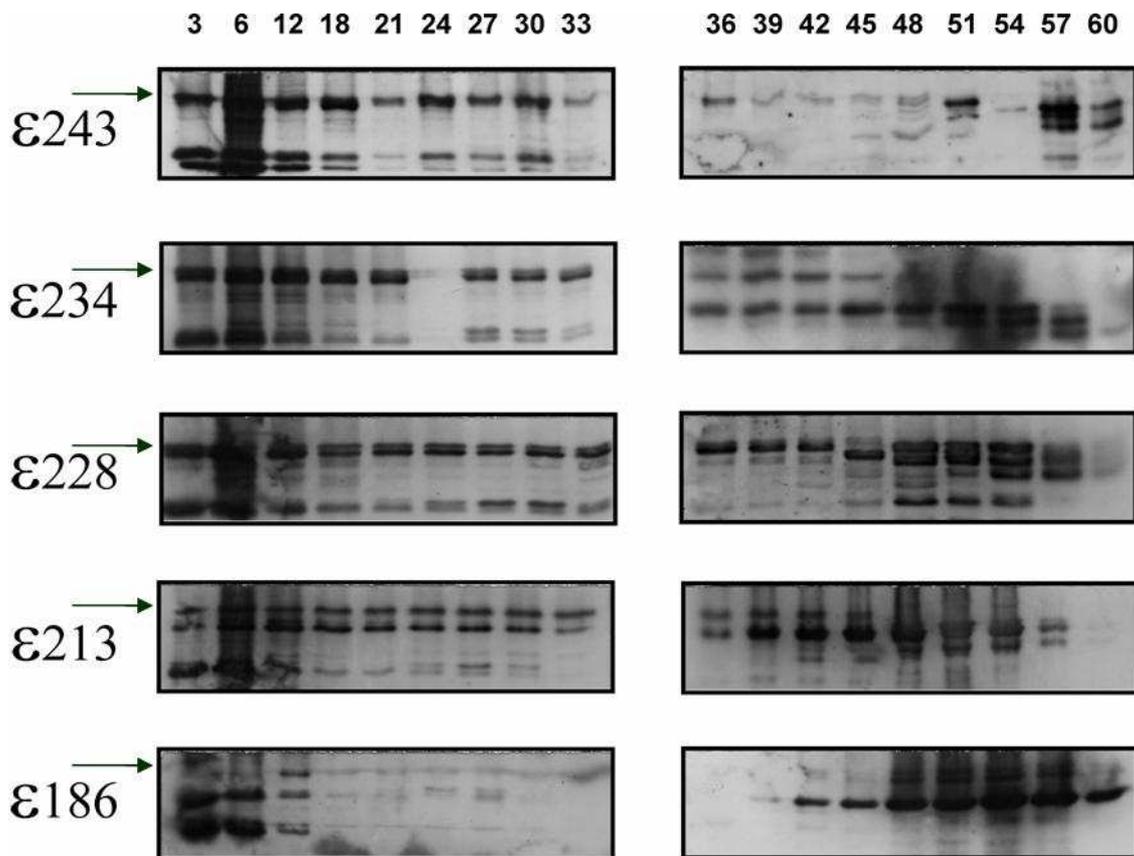


FIG 7: Western blot con anticorpo anti  $\epsilon$  sulle frazioni eluite dalla GF effettuata con estratti proteici ottenuti da cellule TOP10/pBAD che sovraesprimono una forma troncata di  $\epsilon$  (indicata a sinistra). La freccia a sinistra indica la massa molecolare di 27 kDa.

Dai western blotting riportati in figura 7 si può osservare la presenza di  $\epsilon$  anche in altre frazioni oltre a quelle contenenti l'oloenzima (frazione 6) e la forma monomerica (frazione 52). Ad eccezione della variante  $\epsilon_{243}$ , la banda corrispondente ad  $\epsilon$  è visibile più intensamente nelle frazioni intorno alla 42 che contengono cioè proteine con una massa molecolare di circa 60 kDa. Tale evidenza permetterebbe di ipotizzare l'esistenza di un complesso proteico di massa molecolare pari a 60 kDa che contiene  $\epsilon$ . Questo complesso potrebbe essere un omo- o un etero-dimero di  $\epsilon$  che potrebbe rappresentare, per esempio, un intermedio nell'assemblaggio dell'oloenzima. Per ottenere maggiori informazioni sulla composizione di tale dimero sono stati effettuati esperimenti successivi (confronta § 5.2 e 6.2).

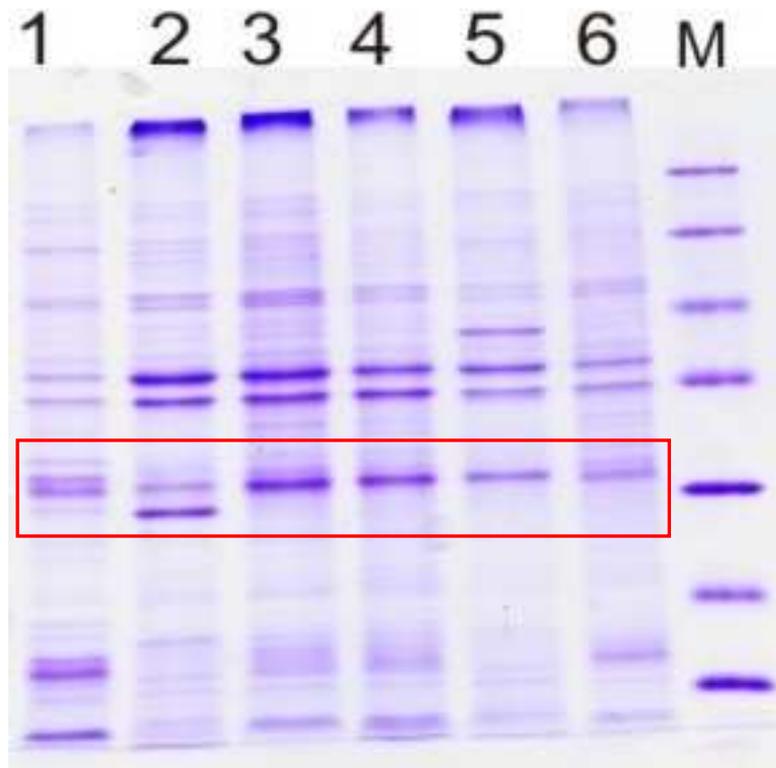


FIG 8: Frazioni 5 di gel filtrazione di TOP 10 contenenti: corsia 1 pBAD- $\epsilon$ 186, 2 pBAD- $\epsilon$ 213, 3 pBAD- $\epsilon$ 228, 4 pBAD- $\epsilon$ 234, 5 pBAD- $\epsilon$ 243, 6 pBAD-AR1, M marker delle seguenti masse molecolari, partendo dal basso 14.4, 18.4, 25, 35, 45, 66.2 e 116 kDa.

Gli esperimenti di western blotting evidenziano chiaramente che la regione C-terminale di  $\epsilon$  rende la proteina fortemente instabile (fig. 9 e 10). In particolare, osservando in figura 9 l'intervallo dalla frazione 52 alla 60 della gel filtrazione della sovraespressione di  $\epsilon$ 243, appare evidente che la forma monomeric di  $\epsilon$  (frazione 52) risulta degradata. Le bande sono infatti deboli e molteplici.

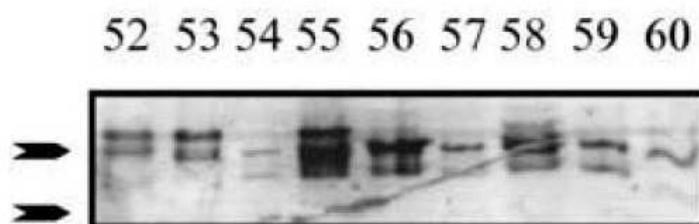


Figura 9: Intervallo dalla frazione 52 alla 60 della gel filtrazione di TOP 10/pBAD-  $\epsilon$ 243. Le frecce indicano la masse molecolari di 25 e 19 kDa.

Dal confronto delle frazioni 52 di figura 10 si evince che la forma più stabile come monomero risulta essere  $\epsilon$  186; il dato per altro conferma i saggi di attività esonucleasica riportati in figura 3. In particolare è possibile notare che se per  $\epsilon$ 186 è presente una sola banda, nel caso delle altre varianti di  $\epsilon$  sono presenti almeno due

bande, elemento che indica che la proteina non è stabile, ma è degradata. Inoltre, pur non trattandosi di un'analisi quantitativa, è possibile affermare che, confrontando l'intensità delle bande di figura 10, questa decresce progressivamente da  $\epsilon$ 186 fino a  $\epsilon$ 243, in accordo con i dati relativi all'attività esonucleasica di  $\epsilon$  e delle sue varianti in forma monomerica.

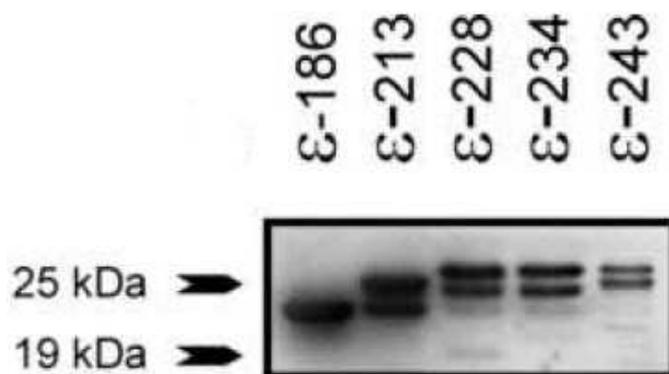


FIG 10: Western blotting delle frazioni 52 a confronto

Dal Western blotting risulta chiaro che la proteina intera, quando si trova in forma monomerica, è soggetta a forte proteolisi, mentre risulta stabile quando si trova protetta in un complesso molecolare di tipo dimerico o associata all'oloenzima.

La instabilità intrinseca di  $\epsilon$  causata dalla regione C-terminale della proteina potrebbe essere il fattore che limita la quantità della subunità esonucleasica nell'assemblaggio dell'oloenzima.

## 2. Impiego di ceppi deleti per proteasi e chaperones molecolari

La subunità  $\epsilon$  della Pol III è instabile e soggetta a degradazione probabilmente ad opera di specifiche proteasi che riconoscono una sequenza aminoacidica specifica a livello della regione C-terminale. Nella cellula però un ruolo chiave per la stabilità delle proteine è svolto anche dagli chaperones molecolari che potrebbero associarsi alla subunità  $\epsilon$  durante alcune fasi del ciclo cellulare preservandola dalle proteasi per poi dissociarsene e lasciare che avvenga la degradazione.

Per approfondire gli studi sulla stabilità e sulla degradazione di  $\epsilon$  in vivo sono stati utilizzati alcuni ceppi di *E. coli* deleti per alcune proteasi o per alcuni chaperones molecolari ottenuti dalla collezione Keio. Per identificare gli enzimi responsabili della degradazione o della stabilità della subunità  $\epsilon$ , è stata costruita una serie di ceppi isogenici di *E. coli* BW25113 che differiscono per l'inattivazione di una singola proteasi o di uno chaperone molecolare. In particolare sono state scelte come bersaglio alcune proteasi citosoliche non

essenziali per la sopravvivenza di *E. coli* quali Lon, ClpP e HslV e alcuni chaperones molecolari (DnaK, Tig, GroL e HtpG).

Sovraesprimendo in tali ceppi la proteina DnaQ risulta possibile ottenere maggiori informazioni su quali sono i fattori molecolari coinvolti nella stabilità o nella degradazione di  $\epsilon 243$ .

Prima di sovraesprimere il costrutto pBAD-*dnaQ* in tali ceppi, è stata effettuata la trasduzione mediata dal fago P1 (figura 11) dell'allele *recA1907::cat-aadA* per evitare la ricombinazione omologa tra il gene *dnaQ* presente nel cromosoma e quello clonato nel vettore pBAD.

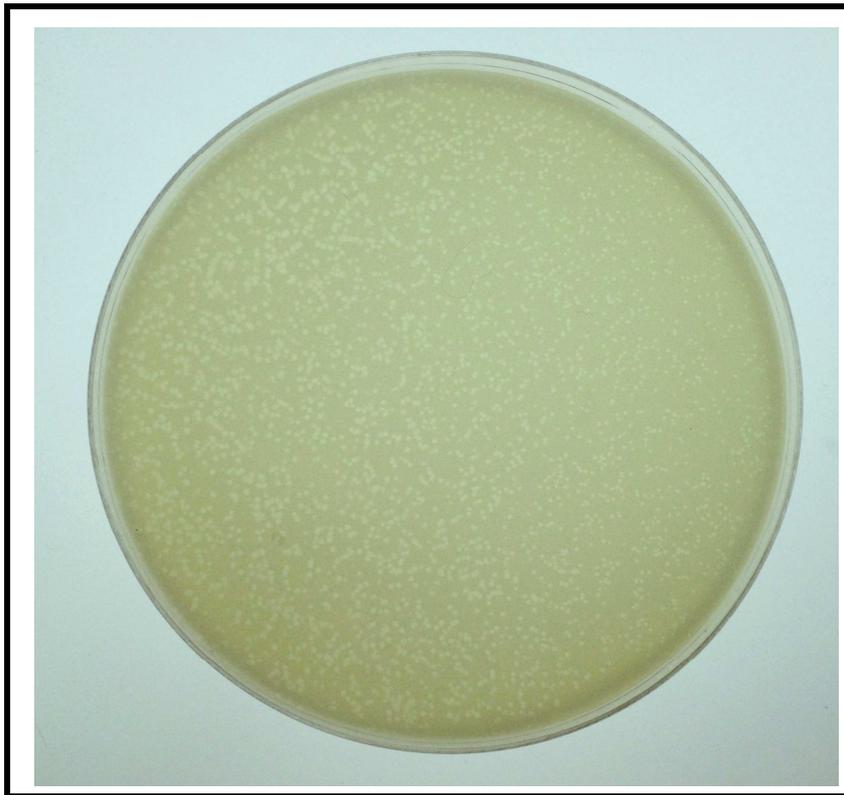


Figura 11: Placche fagiche, preparazione della stock fagica di P1 *vir* contenente l'allele *recA1907::cat-aadA*.

I trasduttanti sono stati trasformati con il costrutto pBAD-*dnaQ* ed il gene *dnaQ* è stato sovraespresso. Sono state estratte le proteine totali e 4 mg di proteina sono state iniettati in colonna cromatografica ad esclusione molecolare.

## 2.1 Analisi Western blot

Per valutare la stabilità di  $\epsilon 243$  in forma monomerica e associata alla Pol III è stata effettuata un'analisi western blotting con l'anticorpo policlonale anti- DnaQ sulle frazioni di massa molecolare corrispondente alla

forma monomerica di  $\epsilon$  (frazione 52) e a quella corrispondente a proteine di massa molecolare compatibile con l'oloenzima (frazione 6).

In figura 12 A appare evidente che nel ceppo *wild type* (WT) di *E. coli* BW25113, progenitore degli altri ceppi utilizzati in questo esperimento,  $\epsilon$ 243 è fortemente instabile poiché appaiono almeno quattro bande di degradazione. La instabilità di  $\epsilon$  risulta essere drammatica in caso di sovraespressione della proteina in ceppi nei quali lo chaperone molecolare DnaK è inattivo.

Un forte decremento della concentrazione cellulare di  $\epsilon$  si osserva anche quando lo chaperone molecolare Tig risulta inattivo. In questo caso sono visibili almeno tre bande di debole intensità. In caso di assenza delle proteasi HslV, Lon o dello chaperone molecolare HtpG si osservano bande meno intense rispetto al *wild type*. Diversamente è possibile rilevare un aumento nella stabilità di  $\epsilon$  quando o la proteasi ClpP o lo chaperone molecolare GroL vengono inattivati: la banda corrispondente alla forma intera di  $\epsilon$  risulta essere più intensa. Pertanto dalla figura 12 A si evince che DnaK, GroL e ClpP sono coinvolte nella stabilità e nella degradazione di  $\epsilon$ .

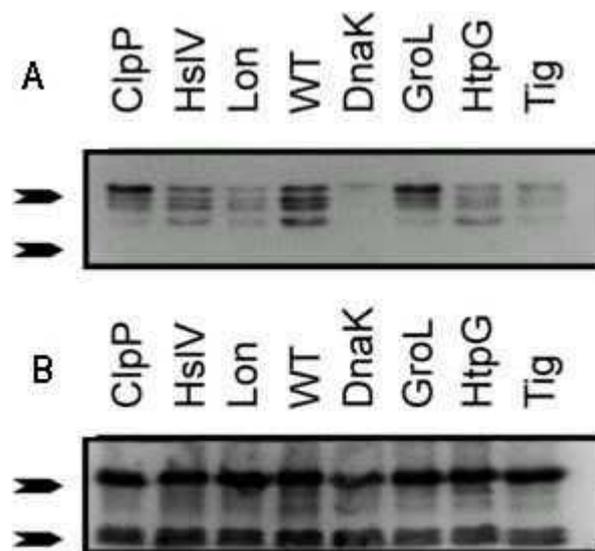


FIG 12: Gel filtrazione di estratti proteici di ceppi deleti per proteasi o per chaperones molecolari (il gene deletato è indicato sopra ogni corsia) nei quali è stata sovraespressa  $\epsilon$ 243. **A** Frazioni 52. **B** frazioni 6. Le frecce indicano i marcatori precolorati di 25 e 19 kDa.

Dall'analisi western blotting effettuata sulle frazioni corrispondenti a proteine di massa molecolare compatibili con l'oloenzima (figura 12B) si evince che la stabilità di  $\epsilon$  è indipendente dal genotipo del ceppo nel quale viene sovraespressa. Infatti l'intensità delle bande riconosciute dall'anticorpo anti  $\epsilon$  non sembra variare nei diversi ceppi deleti rispetto al *wild type*. Confrontando i pannelli A e B della figura 12 si può ipotizzare che se è vero che ClpP, GroL e DnaK sono coinvolti nella stabilità e nella degradazione di  $\epsilon$ , la loro azione molecolare

viene svolta solo fino a quando la proteina  $\epsilon$  si trova in forma monomerica e non influenzano invece la sua stabilità quando questa si trova assemblata nell'oloenzima.

L'analisi western è stata condotta con le frazioni 42 che contengono proteine la cui massa molecolare è compatibile con un dimerico di  $\epsilon$  (figura 13). Anche in questo caso l'assenza dello chaperone molecolare DnaK sembra avere conseguenze più pronunciate rispetto alla delezione degli altri chaperones. Confrontando la figura 12 A con la figura 13 si evince che in quest'ultimo caso non sono presenti bande di degradazione di  $\epsilon$ , dal momento che è presente solamente una banda principale al di sopra dei 25 kDa e sono scomparse invece le bande sottostanti. Questo dato sembra confermare che solo la forma monomerica di  $\epsilon$  è instabile e soggetta a proteolisi. Quando invece si trova protetta all'interno di un complesso proteico, come in un dimerico o nell'oloenzima,  $\epsilon$  è maggiormente stabile.

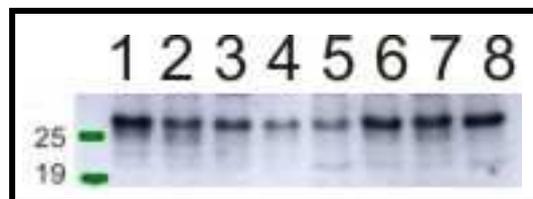


Fig 13: Western blotting delle frazioni 42 della gel filtrazione di estratti proteici di ceppi deleti per proteasi o per chaperones molecolari nei quali è stata sovraespressa  $\epsilon$ 243. I numeri sulle corsie indicano lo specifico gene deletato: 1 CtpP, 2 HslV, 3 Lon, 4 WT, 5 DnaK, 6 GroL, 7 HtpG, 8 Tig. A destra sono indicati i marcatori precolorati di 25 e 19 kDa.

### 3. Realizzazione e impiego del costrutto pBAD-*dnaE*

#### 3.1 Clonaggio di *dnaE* e di sue forme tronche ed analisi fenotipica

Sulla base della struttura secondaria della regione N-terminale, sono state progettate delle forme tronche della subunità  $\alpha$  contenenti una porzione più o meno ampia di tale regione terminale. In particolare, esaminando la struttura terziaria della proteina sono state evidenziate tre  $\alpha$  eliche (residui 288-307, 311-335, 339-353) (Wieczorek e McHenry, 2006). Pertanto sono state sovraesprese forme troncate di  $\alpha$  composte esclusivamente da una porzione più o meno ampia della regione N-terminale e mancanti dei residui catalitici responsabili dell'attività polimerasica (D401, D403, D555). In tale modo è possibile studiare se la forma troncata di  $\alpha$  interagisce ancora con la regione C-terminale di  $\epsilon$ . La sovraespressione della porzione di  $\alpha$  sprovvista di attività polimerasica, interagendo con  $\epsilon$ , priva la cellula della subunità con attività esonucleasica.

La replicazione da parte della Pol III naturalmente espressa, essendo priva di  $\epsilon$ , risulterà meno accurata e meno fedele e il fenotipo risultante sarà di tipo mutatore.

Pertanto sono stati realizzati i costrutti (figura 14):

- pBAD- $\alpha$ 287 (privo delle tre  $\alpha$  eliche)
- pBAD- $\alpha$ 308 (fornito di una sola  $\alpha$  elica)
- pBAD- $\alpha$ 336 (fornito di due delle tre  $\alpha$  eliche)
- pBAD- $\alpha$ 353 (fornito delle tre  $\alpha$  eliche)

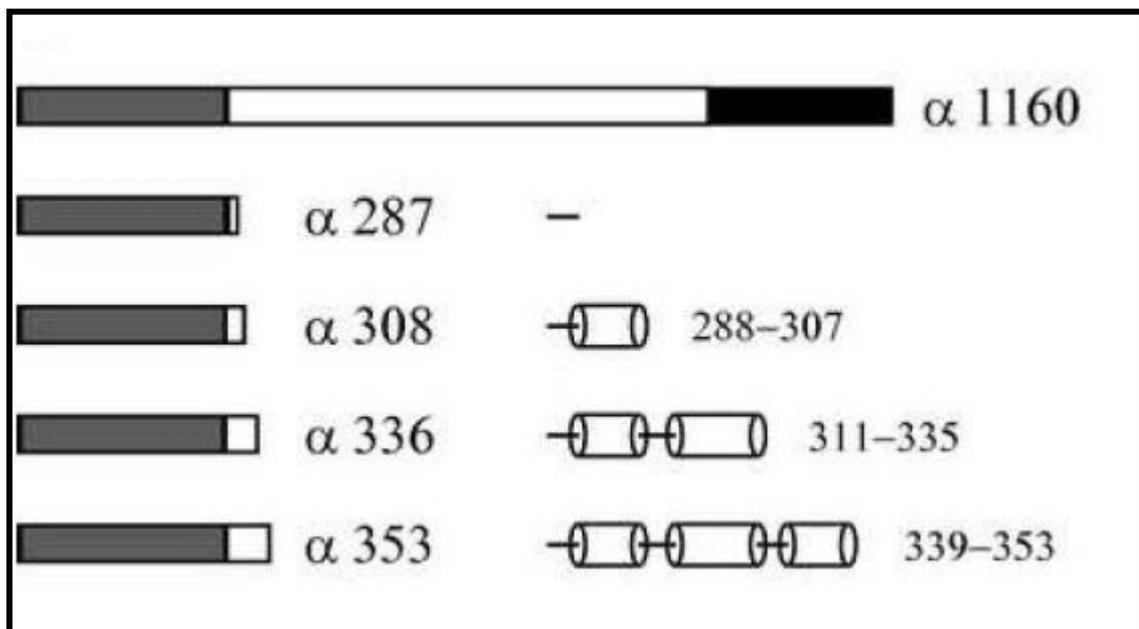


FIG 14 :  $\alpha$  intera e sue forme tronche. Il dominio PHP N-terminale è in grigio. Alla destra di ogni variante di  $\alpha$  sono indicate le  $\alpha$  eliche presenti.

Tali costrutti contenenti  $\alpha$  intera o sue forme tronche sono stati utilizzati per trasformare ceppi di *E.coli* BW25993 elettrocompetenti. In tali ceppi era stato precedentemente trasdotto l'allele *recA1907::cat-aadA* per evitare la ricombinazione omologa tra la copia di *dnaE* presente nel cromosoma e quella clonata nel plasmide. La espressione della subunità  $\alpha$  o delle sue forme tronche è stata indotta con arabinosio in terreno solido per 30-50 ore; contemporaneamente, è stata indotta l'espressione dell'operone *lac* mediante isopropil- $\beta$ -D-tiogalattopiranoside (IPTG). Fa parte di tale operone il gene *lacZ* che codifica per la  $\beta$ -galattosidasi, un enzima capace di degradare composti di natura  $\beta$ -galattosidica nei monomeri costituenti; dal lattosio, ad esempio, si liberano glucosio e galattosio. Il substrato cromogenico sintetico 5-bromo-4-cloro-3-indolil- $\beta$ -D-galattopiranoside (X-Gal) (figura 15 A) è un analogo del lattosio e conferisce colorazione blu alle colonie che

esprimono  $\beta$ -galattosidasi; pertanto, tale substrato può essere utilizzato come indicatore qualitativo dell'attività  $\beta$ -galattosidasica.

In un potenziale ceppo mutatore la frequenza di mutazioni spontanee a carico del genoma aumenta rispetto a quanto accade nel corrispondente ceppo *wild-type*. Tali mutazioni avvengono casualmente in tutto il genoma, ma non sempre possono essere rilevabili fenotipicamente. Se tali mutazioni dovessero avvenire nel gene codificante per la  $\beta$ -galattosidasi potrebbero però essere identificate. Pertanto *lacZ* è stato scelto come gene reporter di mutazioni che possono insorgere qualora la sovraespressione della subunità  $\alpha$  o di una sua parte sequestri  $\epsilon$ , generando un fenotipo mutatore.

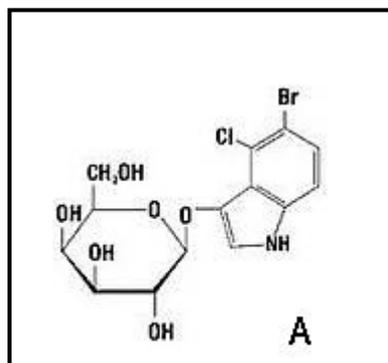
In figura 15 (B-G) si può osservare l'effetto fenotipico determinato dalla sovraespressione delle varianti di  $\alpha$ .

In particolare, per alcune forme troncate di  $\alpha$ , si possono distinguere sulle colonie blu che esprimono una  $\beta$ -galattosidasi *wild type*, in grado cioè di degradare il substrato X-gal, delle papille bianche (E, F, G). Sovraesprimendo  $\alpha$ 336 (E),  $\alpha$ 353 (F) e  $\alpha$  intera (G) le colonie appaiono papillate, fenotipo non presente nel controllo (B) né quando  $\alpha$ 287 (C) o  $\alpha$ 308 (D) vengono sovraesprese.

Nel gene *lacZ* di queste colonie papillate sono insorte una o più mutazioni che non rendono più in grado la  $\beta$ -galattosidasi di degradare il substrato X-gal.

Analizzando i risultati di figura 15 si può affermare che  $\alpha$ 287 e  $\alpha$ 308 non sono in grado di legare  $\epsilon$ . Infatti quando vengono sovraesprese, il fenotipo risultante non è distinguibile da quello del controllo negativo. Diversamente, sovraesprimendo le varianti  $\alpha$ 336,  $\alpha$ 353 ed  $\alpha$  intera il fenotipo risulta mutatore.

Pertanto è possibile ipotizzare che le  $\alpha$  eliche E311-M335 e G339-D353 siano determinanti nel legame ad  $\epsilon$ . Wieczorek and McHenry (2006) avevano dimostrato che la regione di  $\alpha$ , determinante per il legame ad  $\epsilon$ , comprendeva i primi 320 aminoacidi della proteina. Le osservazioni fenotipiche mostrate in figura 14, estendono la regione N-terminale di  $\alpha$  che interagisce con  $\epsilon$  fino almeno al residuo 353.



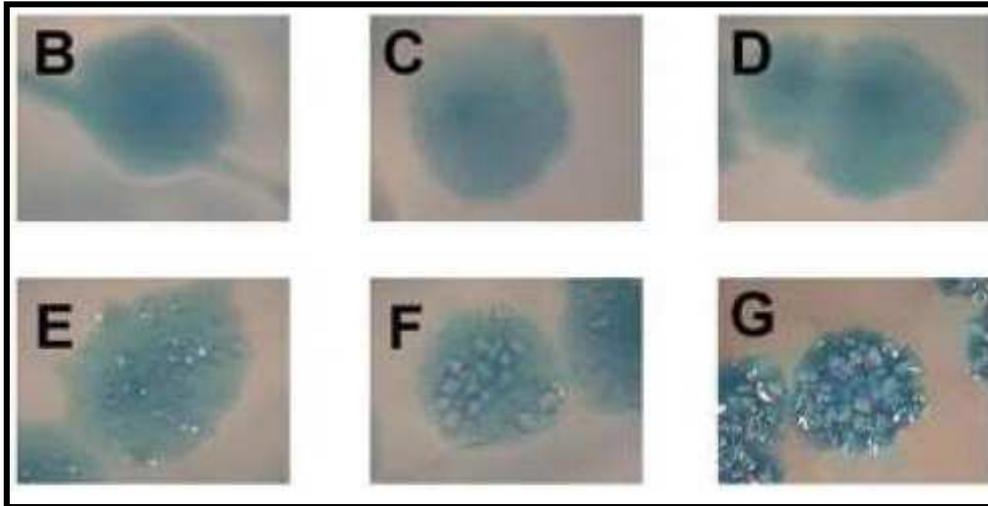


FIG15 : **A** X-gal, il substrato cromogenico della  $\beta$ -galattosidasi. **B-G**: colonie di *E. coli* BW25993 trasformate con pBAD (**B**), o con lo stesso vettore contenente  $\alpha$ 287 (**C**),  $\alpha$ 308 (**D**),  $\alpha$ 336 (**E**),  $\alpha$ 353 (**F**),  $\alpha$ 1160 (**G**) e coltivate in piastra contenente LB , IPTG, arabinosio e X-Gal.

#### 4. Realizzazione e impiego del costrutto pBAD-His-dnaQ

Per approfondire le informazioni relative alle forme di degradazione di  $\epsilon$  che si generano in vivo, il gene *dnaQ* è stato clonato con una regione codificante 36 aminoacidi aggiuntivi tra cui sei istidine (massa molecolare 3989 Da) situate all'N-terminale. Il gene His-*dnaQ* è stato sovraespresso (massa molecolare 31 kDa). Le proteine totali sono state estratte e caricate in una colonna cromatografica di affinità. Le frazioni contenenti attività esonucleasica sono state unite e caricate in colonna ad esclusione molecolare. Le frazioni eluite da tale colonna contengono  $\epsilon$  legata al motivo esaistidinico e quelle della frazione 42 (massa molecolare di circa 60 kDa) sono state separate attraverso SDS-PAGE (figura 16).

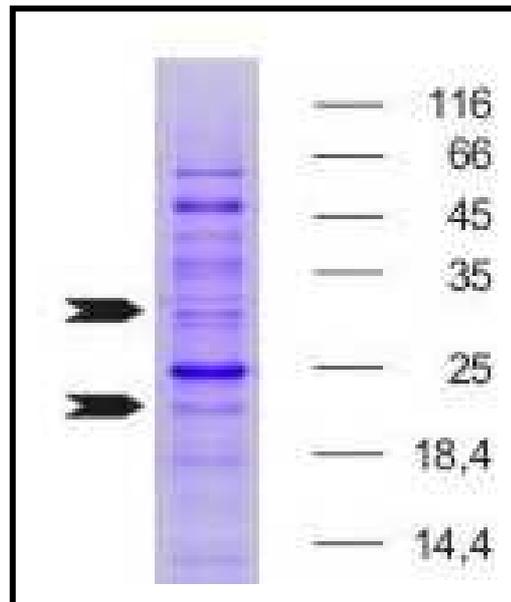


Fig 16: SDS-PAGE della frazione 42 eluita dalla gel filtrazione effettuata dopo la purificazione d'affinità di cellule TOP10/pBAD-His-*dnaQ*. Le frecce indicano la posizione delle bande escisse.

Le due frecce in figura 16 indicano la posizione delle due bande escisse dal gel elettroforetico. La massa molecolare delle proteine presenti in tali bande è stata calcolata utilizzando la curva di calibrazione della figura 17. In tale curva sono riportati i valori di mobilità elettroforetica relativa ( $R_f$ ) di proteine di massa molecolare nota caricate nel medesimo gel elettroforetico. Utilizzando tale curva di calibrazione, le masse delle proteine escisse risultano essere rispettivamente 31 e 22.6 kDa.

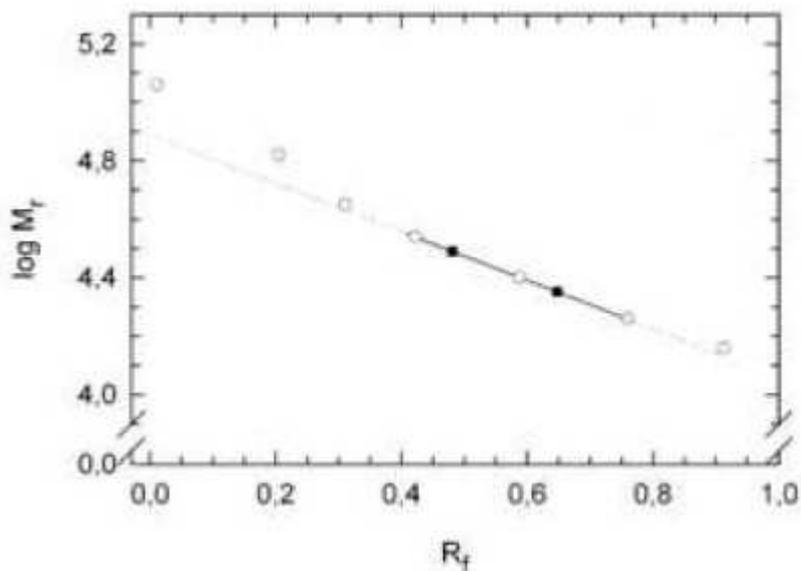


FIG 17: Curva di calibrazione ottenuta utilizzando la mobilità elettroforetica relativa ( $R_f$ ) di proteine (○) di massa molecolare nota ( $\log M_r$ ). La regione lineare di tale curva ha permesso di determinare la massa molecolare delle due bande escisse dal gel SDS-PAGE (●).

## 4.1 Analisi attraverso spettrometria di massa

Le proteine di 31 e 22.6 kDa, indicate dalle frecce in figura 16, sono state escisse dal gel SDS-PAGE e digerite con tripsina. I frammenti ottenuti sono stati analizzate attraverso spettrometria di massa (Q-TOF).

Nella tabella sottostante sono elencati i peptidi identificati in seguito a Q-TOF. I numeri tra parentesi in grassetto si riferiscono alla posizione che gli stessi aminoacidi occupano nella proteina  $\epsilon$  priva della coda di 36 aminoacidi aggiuntivi all'N-terminale.

<b>Banda di 31 kDa (compatibile con la forma intera di <math>\epsilon</math>)</b>			<b>Banda di 22,6 kDa (compatibile con una forma tronca di <math>\epsilon</math>)</b>		
Peptide determinato	Peso molecolare teorico (Da)	Peso molecolare sperimentale (Da)	Peptide determinato	Peso molecolare teorico (Da)	Peso molecolare sperimentale (Da)
37-43 <b>(1-7)</b>	748.39	748.39	37-43 <b>(1-7)</b>	748.39	748.39
66-77 <b>(30-41)</b>	1310.76	1310.74	66-77 <b>(30-41)</b>	1310.76	1310.75
90-112 <b>(54-76)</b>	2552.29	2552.27	90-112 <b>(54-76)</b>	2552.29	2552.26
93-112 <b>(57-76)</b>	2184.11	2184.08	93-112 <b>(57-76)</b>	2184.11	2184.10
163-171 <b>(127-135)</b>	930.51	930.51	128-151 <b>(92-115)</b>	2699.33	2699.29
172-178 <b>(136-142)</b>	862.48	862.47	163-171 <b>(127-135)</b>	930.51	930.50
244-264 <b>(208-228)</b>	2340.22	2340.21	172-178 <b>(136-142)</b>	862.48	862.46
			241-249 <b>(205-213)</b>	1069.67	1069.60

L'analisi della proteina di 22.6 kDa attraverso Q-TOF permette di identificare alcuni dei peptidi che si generano. Tali dati possono fornire informazioni sulle forme di  $\epsilon$  che si generano, in vivo, in seguito a proteolisi.

Nella banda di 22.6 kDa sono stati identificati tutti i residui da M1 a R213 che hanno una massa complessiva di 23,773 Da. Dal momento che la massa molecolare determinata in seguito alla corsa elettroforetica è di 22.6 kDa (figura 17), è possibile ipotizzare che il principale prodotto di proteolisi in vivo sia  $\epsilon$ 213.

## 5. Utilizzo del costrutto antisense pBAD-AR1

La proteolisi della regione C-terminale di  $\epsilon$  origina delle forme tronche non più in grado di associarsi con  $\alpha$ . Pertanto la stabilità di  $\epsilon$  rappresenta il fattore limitante per l'assemblaggio di Pol III cataliticamente attiva.

I dati sperimentali fin qui descritti sono stati ottenuti sovraesprimendo  $\epsilon$  intera o sue varianti tronche. Per ottenere maggiori informazioni sul ruolo di  $\epsilon$  nell'assemblaggio dell'oloenzima, è stato analizzato il fenotipo correlato al livello di espressione naturale di  $\epsilon$ .

Per tale scopo è stato utilizzato un RNA antisense (AR1) anti-*dnaQ* la cui capacità di silencing era già stata descritta (Stefan et al., 2003). Il costrutto pBAD-AR1 contiene, nel vettore di espressione pBAD, un frammento di *dnaQ* inserito con orientamento invertito. L'espressione dell' RNA antisense può essere indotta con arabinosio; se l'antisense AR1 viene espresso per 12 ore, si osserva un forte decremento dell' attività proofreading nell' estratto proteico totale (Tesi di laurea di Stefano Cianchetta "Studi funzionali mediante strategia antisense della subunità  $\epsilon$  della DNA polimerasi III di *Escherichia coli*, anno accademico 2002-2003) come si evince dalla tabella sottostante.

	NON INDOTTO	INDOTTO
<b>Attività</b> (nM / s · mgP)	98.0 ± 9.5	46.1 ± 0.6
<b>Attività</b> (aM / s · cell)	650.1 ± 240.5	98.0 ± 25.7
<b>Concentrazione di proteina</b> (fg · cell)	190	76

### 5.1 Gel filtrazione e saggio di attività esonucleasica

Stefan e collaboratori hanno osservato un rallentamento della crescita cellulare in caso di induzione dell' RNA antisense AR1 (Stefan et al., 2003). Per ottenere maggiori informazioni su questo fenotipo, cellule TOP10 di

*Escherichia coli* sono state coltivate per 12 ore a 37 °C inducendo con arabinosio 13 mM o non inducendo l'espressione di AR1. 4 mg delle proteine totali estratte sono state caricate in colonna cromatografica ad esclusione molecolare. Con le frazioni corrispondenti a proteine di massa molecolare compatibili con l'oloenzima è stato effettuato un saggio di attività esonucleasica (fig 18)

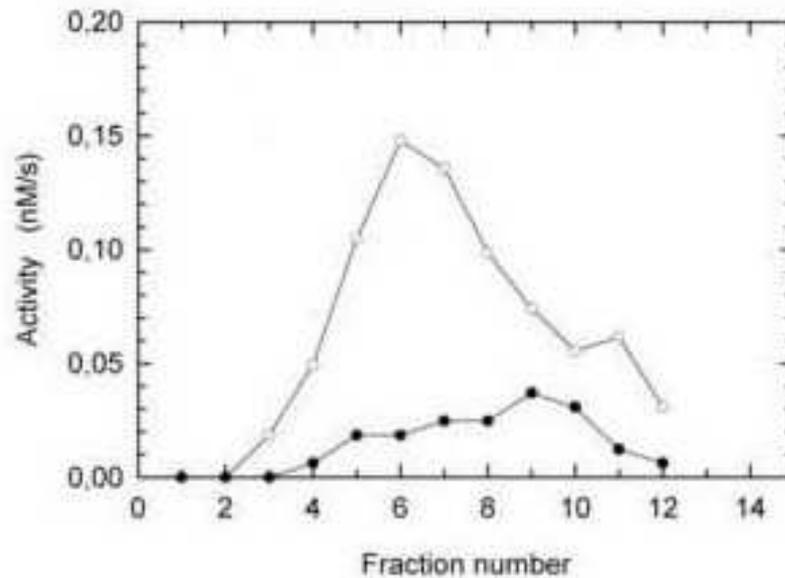


FIG 18 Attività esonucleasica delle frazioni 0-12 eluite dalla gel filtrazione di proteine estratte da *E. coli* TOP10/pBAD-AR1 non indotto (○) o indotto(●)

Come si osserva in figura 18, l'espressione di AR1 determina un decremento nei livelli di attività dell'oloenzima. Tale decremento potrebbe essere la causa del ritardo nella crescita cellulare che era stato osservato (Stefan et al., 2003).

L'attività esonucleasica determinata nella frazione contenente l'oloenzima di estratti cellulari di *E. coli* TOP 10/pBAD-AR1 non indotto (figura18, valore di 0.15 nM/s) è assolutamente confrontabile con l'attività delle frazioni dell'oloenzima misurata nel controllo negativo *E. coli* TOP 10/pBAD (0.13 nM/s) riportata in figura 19. Pertanto questo valore di attività esonucleasica riflette il numero di copie di oloenzima Pol III presenti nella cellula in caso di normale espressione di *dnaQ*.

Il decremento di attività esonucleasica che si verifica in caso di espressione di AR1 potrebbe essere un riflesso della diminuzione del numero di copie di oloenzima per cellula rispetto alle condizioni di normale espressione di  $\epsilon$ .

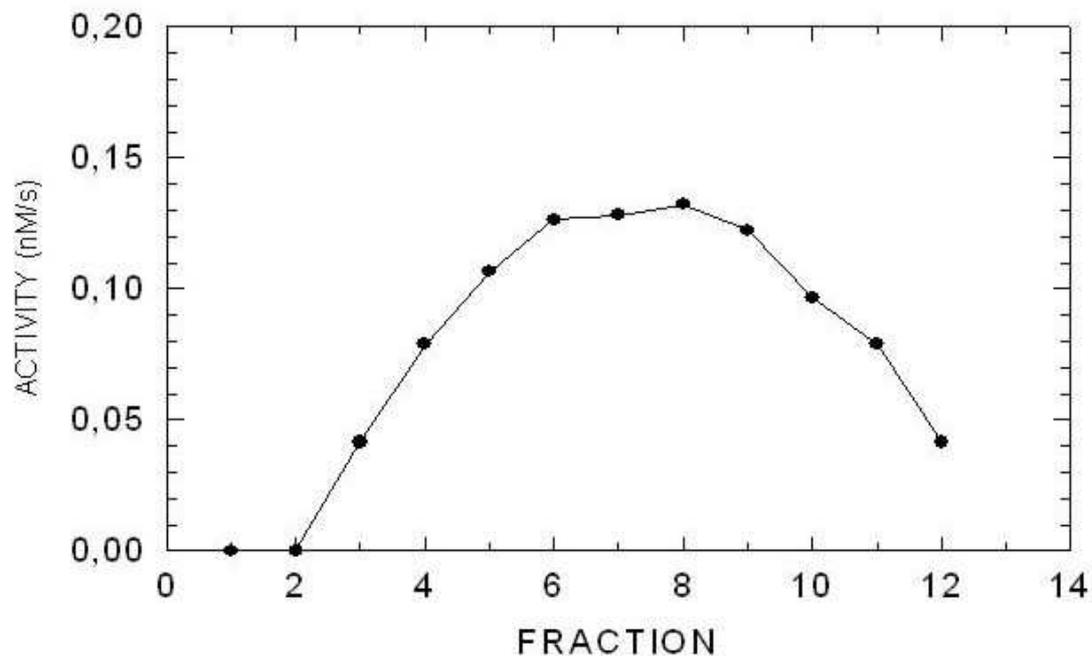


FIG 19 Attività esonucleasica delle frazioni 0-12 eluite dalla gel filtrazione di proteine estratte da *E. coli* TOP 10/pBAD.

I saggi di attività esonucleasica, effettuati sulle frazioni della gel filtrazione dei campioni nei quali l'antisense AR1 è stato indotto e nei quali non è stato indotto, hanno permesso di evidenziare due picchi di attività esonucleasica corrispondenti a proteine con massa molecolare pari a 60 e 30 kDa (fig 20).

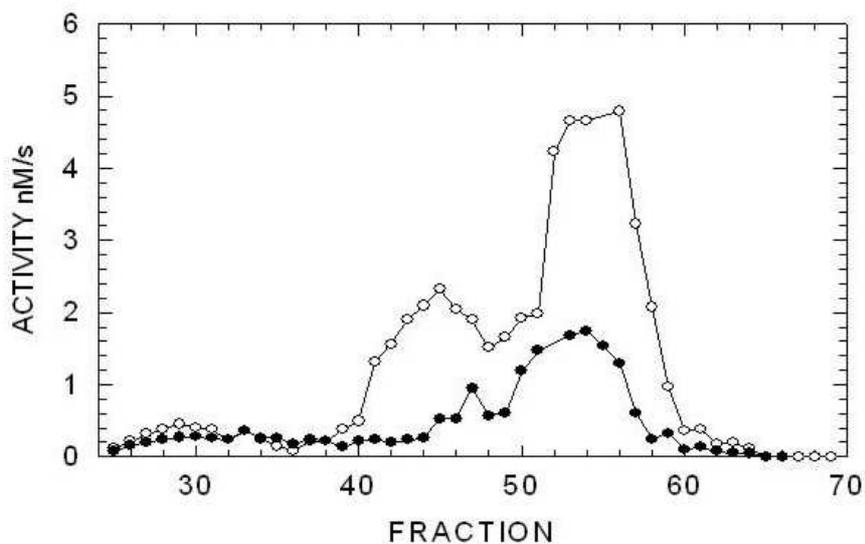


FIG 20 Attività esonucleasica delle frazioni 25-70 eluite dalla gel filtrazione di proteine estratte da *E. coli* TOP10/pBAD-AR1 non indotto (○) o indotto(●)

Il picco di attività di 30 kDa rappresenta la subunità  $\epsilon$  della DNA polimerasi III. L'induzione dell'antisense AR1 causa una riduzione di attività esonucleasica del 70%.

Il picco di attività corrispondente ad una proteina di massa molecolare di circa 60 kDa potrebbe rappresentare un complesso proteico contenente, tra le altre, la subunità  $\epsilon$ . Possono essere fatte delle ipotesi circa questo complesso molecolare. Potrebbe trattarsi di un dimero di  $\epsilon$  oppure della subunità  $\epsilon$  complessata a DNA o ad altre proteine. La presenza di  $\epsilon$  in tali frazioni era stato osservato anche mediante analisi western blot (figure 7 e 13). Il picco corrispondente alla frazione di massa molecolare pari a circa 60 kDa ha un'attività inferiore rispetto a quello in corrispondenza di frazioni di massa di circa 30 kDa.

Le proteine eluite nella frazione 42 del campione nel quale non è stato indotto AR1, sono state precipitate in acido tricloroacetico e caricate in SDS-PAGE (fig 21).

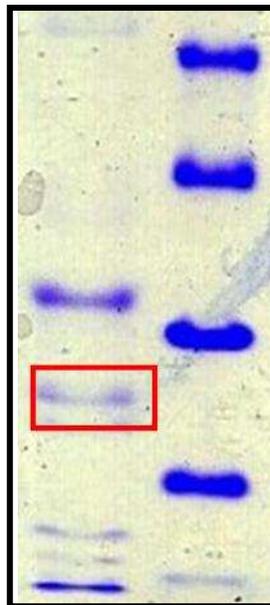


FIG 21: SDS-PAGE della frazione 42 della gel filtrazione TOP 10/pBAD-AR1 non indotto. In rosso la banda escissa. Nella corsia di sinistra sono stati caricati i marker di massa molecolare pari a 25, 35, 45 e 66.2 kDa

## 5.2 Analisi di massa

Per ottenere maggiori informazioni riguardo la presenza di  $\epsilon$  nelle frazioni corrispondenti a proteine di circa 60 kDa è stata effettuata un'analisi di massa che ha permesso di evidenziare quattro frammenti della regione N-terminale, come indicato nella tabella sottostante.

Peptide determinato di DnaQ	Peso molecolare teorico (Da)	Peso molecolare sperimentale (Da)
<b>152-158</b>	867,397	867,403
<b>127-135</b>	930,513	930,542
<b>30-41</b>	1310,756	1310,789
<b>43-56</b>	1672,868	1672,868

L'analisi conferma la presenza della subunità  $\epsilon$  nelle frazioni contenenti proteine la cui massa è pari a circa 60 kDa. Pertanto è possibile ipotizzare che la subunità esonucleasica sia contenuta in un complesso molecolare di 60 kDa.

### 5.3 Numero di copie di oloenzima Pol III

Dopo aver determinato il contenuto di proteine per cellula e l'attività di proofreading del picco cromatografico contenente l'oloenzima, considerando il valore di  $k_{cat}$  per  $\epsilon$  quando si utilizza 5'-p-nitrofenil estere della timidina 5'-monofosfato come substrato della reazione esonucleasica, (Hamdan et al., 2002) è possibile calcolare il numero di copie di oloenzima di DNA Pol III per cellula. Il numero di copie di oloenzima per cellula in condizioni normali è di 12-14 molecole, valore in accordo con quanto riportato in letteratura (Maki et al., 1985 e Maki e Kornberg, 1985) mentre in caso di espressione dell'RNA antisenso AR1 il numero di molecole di oloenzima scende ad 1 copia per cellula.

## 6. Analisi dell'estratto proteico di cellule TOP10

### 6.1 Coltura, lisi e gel filtrazione

Per ottenere maggiori informazioni sul comportamento in vivo della subunità  $\epsilon$  e sulle forme di aggregazione presenti nella cellula, è stata esaminata la presenza di tale proteina in *Escherichia coli* TOP10 senza ricorrere alla sua sovraespressione. In particolare si è scelto di analizzare la natura del dimero di  $\epsilon$  che era stato identificato nell'ambito di esperimenti precedenti (cfr § 5.2 ). Pertanto sono state coltivate cellule TOP10 e sono state estratte le proteine totali. Il pool di proteine cellulari è stato separato con una colonna ad esclusione

molecolare. Le proteine eluite nelle frazioni con una massa molecolare pari a due volte la massa della subunità  $\epsilon$  (frazione  $42 \pm 1$ ) sono state separate attraverso gel bidimensionale.

## 6.2 Gel bidimensionale e analisi attraverso spettrometria di massa

Per ottenere maggiori informazioni riguardo alla composizione del dimero (omo- o eterodimero) sono stati analizzati alcuni degli spots del gel bidimensionale. Se è vero, come è stato precedentemente dimostrato, che nelle frazioni contenenti proteine la cui massa molecolare è circa 60 kDa è presente  $\epsilon$ , si tratta di verificare se, insieme a questa proteina di circa 27 kDa, è presente un'altra proteina con una massa molecolare che, sommandosi a quella della subunità esonucleasica, dia un dimero di massa molecolare complessiva pari a circa 60 kDa. Qualora si escludesse la presenza di una proteina con queste caratteristiche, si potrebbe concludere che  $\epsilon$  si trova in un complesso proteico omodimerico. Viceversa se le proteine identificate avessero le caratteristiche ricercate, si potrebbe ipotizzare la presenza di un eterodimero contenente anche  $\epsilon$ . La subunità  $\epsilon$  ha una massa molecolare di 26.835 Da e un punto isoelettrico di 5.65. Sono stati analizzati attraverso spettrometria di massa alcuni spots con massa molecolare pari a circa 30 kDa proprio perché, se avvenisse la formazione di un complesso con  $\epsilon$ , il complesso molecolare finale sarebbe di circa 60 kDa quindi compatibile con la massa della frazione isolata in gel filtrazione e analizzata nel gel bidimensionale. I sette spots scelti per l'analisi di massa sono indicati nella figura 22 con un cerchietto rosso.

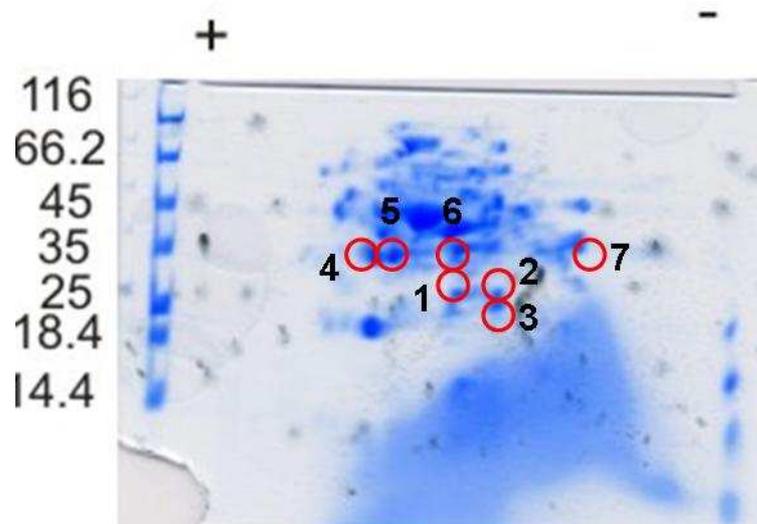


Fig 22: Gel bidimensionale delle frazioni 41, 42, 43 della gel filtrazione di cellule TOP10. I cerchietti in rosso con il rispettivo numero indicano gli spots escissi dal gel. A sinistra è stato caricato il marker. Sopra al gel è indicata la posizione dell'anodo e del catodo utilizzati per la focalizzazione in base al punto isoelettrico della prima dimensione.

Nella tabella sottostante sono indicati invece i risultati dell'analisi di massa di cinque dei sette spots analizzati. Per ogni spot sono indicati i frammenti che sono stati rilevati e, in fondo alla tabella, la proteina putativa di *Escherichia coli* che può avere dato origine a tali frammenti.

SPOT 1	SPOT 2	SPOTS 4 e 5		SPOT 6	SPOT 7
1106,56	590,3621	542,39	1030,58	628,833	1301,713
1108,49	634,3975	578,29	1034,45	643,4	1328,725
1301,74	658,3804	612,31	1084,58	659,34	873,4956
1600,77	738,3802	626,38	1100,65	671,4	737,9039
1790,72	770,3541	667,36	1141,71	704,36	2986,439
1975,95	1004,323	691,37	1251,54	729,45	1819,858
2134,02	1036,497	724,39	1458,72	790,48	1762,873
2551,24	1419,763	739,44	1474,72	806,47	1637,862
	2314,383	740,5	1492,69	847,5	
	2778,89	776,38	1492,71	889,48	
		836,45	1515,73	898,5	
		846,45	2100,02	966,47	
		846,5	2340,02	1007,55	
		889,45	2357,24	1059,64	
		932,54	2406,13	1073,6	
		938,63	2414,24	1148,67	
		979,59	2519,27	1205,69	
		995,66	2715,39	1246,74	
<b>COMPATIBILE CON DNAQ (~27 kDa)</b>	<b>COMPATIBILE CON MANNOSIO PERMEASI (~64 kDa)</b>	<b>COMPATIBILE CON TRANSALDOLASI B (~35 kDa)</b>		<b>COMPATIBILE CON MALATO DEIDROGENASI (~32 kDa)</b>	<b>COMPATIBILE CON GAMMA INTIMINA (~kDa)</b>

Lo spot 3 non è stato identificato, mentre i frammenti che si generano in seguito ad analisi di massa dello spot 5 sono compatibili con la proteina transaldolasi B modificata post-traduzionalmente.

La mannosio permeasi, identificata nello spot 2, è una proteina di massa molecolare di circa 64 kDa che in *E. coli* si trova nella membrana cellulare. Essendo invece  $\epsilon$  una proteina citosolica è possibile escludere che formi un dimero con  $\epsilon$ .

Nello spot 4 è stata identificata la proteina transaldolasi B. L'enzima si trova nel citosol, ha una massa molecolare di 35 kDa e fa parte della via metabolica dei pentosi fosfati. Le reazioni non ossidative della via dei pentosi fosfati consistono nella interconversione degli zuccheri fosforilati a 3, 6 e 5 atomi di carbonio finalizzata alla produzione di riboso- 5- fosfato utilizzato per la sintesi dei nucleotidi.

Nello spot 6 è stato identificato l'enzima malato deidrogenasi, enzima del ciclo di Krebs di massa molecolare di 32 kDa che catalizza la reazione di conversione del malato in ossalacetato.

Infine nello spot 7 è stata identificata la proteina gamma intimina che è una adesina batterica. La proteina viene espressa sulla superficie del batterio per favorire l'adesione alla cellula ospite. Trovandosi in

compartimenti cellulari diversi, si può escludere che si formi un eterodimero tra la gamma intimo e la subunità  $\epsilon$  della Pol III.

### 6.3 Presenza di omodimeri di DnaQ

Non si può escludere la formazione di un eterodimero di malato deidrogenasi o di transaldolasi B con la subunità esonucleasica della DNA polimerasi III perché le masse molecolari e la compartimentazione cellulare risultano compatibili con la massa delle proteine separate nel gel bidimensionale. I ruoli funzionali della malato deidrogenasi, della transaldolasi e di DnaQ sono però molto diversi, quindi con alta probabilità si può affermare che nella cellula si formino in vivo dei complessi omodimerici di  $\epsilon$ .

Tali complessi omodimerici potrebbero conferire stabilità ad  $\epsilon$ . La formazione di un complesso omodimerico di  $\epsilon$  potrebbe essere un passaggio intermedio verso l'assemblaggio della DNA polimerasi III.

## 7. Curve di crescita di BW25113 wild type e $\Delta dnaK734::kan$

In letteratura è noto che le mutazioni nel gene *dnaK* sono permissive a 30 °C, ma che la sopravvivenza della cellula non è garantita quando la temperatura raggiunge i 42 °C (Paek and Walzer, 1987).

Per cercare di trovare una correlazione tra i livelli cellulari di  $\epsilon$  e il fenotipo di crescita sono state confrontate le velocità di crescita di colture di *E. coli* BW25113 *wild type* e del suo ceppo isogenico non in grado di produrre la proteina DnaK funzionale. Dopo 5 ore di crescita a 30 °C, le cellule sono state trasferite ad una temperatura di 42 °C (figura 23). In queste condizioni il ceppo  $\Delta dnaK734::kan$  arresta la sua crescita, mentre il ceppo *wild type* continua a replicarsi.

Quando viene sovraespresso il gene *dnaQ*, il ceppo  $\Delta dnaK734::kan$  continua a crescere anche quando la temperatura di allevamento è pari a 42 °C.

Da questo esperimento è possibile dedurre che DnaK controlla la quantità di  $\epsilon$  in grado di associarsi ad  $\alpha$  ed è essenziale per la crescita a 42 °C.  $\epsilon$  è il fattore limitante per l'assemblaggio del core catalitico della DNA polimerasi III.

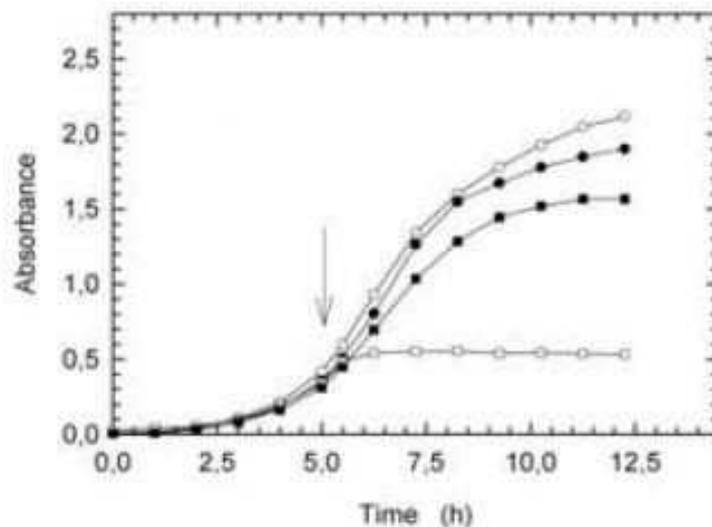


Fig 23: Cinetica di crescita di *E. coli* BW25113 wild type (○) e contenente la proteina DnaK inattiva (□) trasformati con il vettore vuoto pBAD o con il costrutto pBAD- $\epsilon$ 243 (●, ■). Il terreno di crescita è LB contenente 0.05 % NaCl e arabinosio 1 mM. La temperatura di crescita è 30 °C e al tempo indicato dalla freccia, la temperatura è stata alzata a 42 °C.

## 7.1 Fattori molecolari coinvolti nella stabilità di $\epsilon$

La regione C-terminale della subunità  $\epsilon$  è sensibile alla proteolisi ed  $\epsilon$ 213 è la forma di degradazione della subunità esonucleasica della Pol III che probabilmente si genera in vivo. Inoltre sono stati identificati i fattori molecolari coinvolti nella degradazione di  $\epsilon$ . In particolare, mentre ClpP e GroL sono responsabili della degradazione di  $\epsilon$ , DnaK è invece essenziale per la sua stabilità. L'associazione di  $\epsilon$  ad  $\alpha$  è in grado di conferire stabilità alla subunità esonucleasica, pertanto la disponibilità di  $\alpha$  e la quantità di  $\epsilon$  in grado di legarsi ad  $\alpha$ , condizionano i livelli del core catalitico della DNA polimerasi III.

L'inattivazione dello chaperone molecolare DnaK determina un fenotipo non vitale a 42 °C e la sovraespressione di  $\epsilon$  annulla questo effetto. Pertanto DnaK potrebbe assistere il giusto ripiegamento di  $\epsilon$  proteggendola dalla degradazione della regione C-terminale.

Le funzioni molecolari svolte da DnaK e da Tig sono parzialmente sovrapposte e in letteratura è noto che un doppio mutante *dnaK tig* non è vitale e non può essere realizzato (Deuerling et al., 1999). Questa parziale sovrapposizione spiega perché una singola inattivazione di *dnaK* o di *tig* porta ad un decremento della proteina  $\epsilon$ , ma non determina un decremento nel livello dell'olenzima (figura 12 A e B). Dal momento che un mutante nel gene *tig* a 42 °C è vitale, la sopravvivenza della cellula a questa temperatura dipende da DnaK. Dal momento che la sovraespressione di *dnaQ* a 42 °C complementa il genotipo deficiente nel locus *dnaK*, è possibile ipotizzare che la subunità  $\epsilon$  rappresenti il fattore limitante per l'assemblaggio della DNA polimerasi III.





*V. cholerae* DFLVMKSRGEGLEERLEFLFPDPEVRAKRRPEYDERLQVELDVINQMGPFGYFLIVMEFIQ 333  
*V. harveyi* DFLVMKSRGEGLEERLEFLFPDPEERKQRRPEYDRLQIELDVINQMGPFGYFLIVMEFIQ 333  
*V. fischeri* DFLVVKSQEGLEERLAFLEFPDEKVRLERRPEYDRLQIELDVINQMGPFGYFLIVMEFIQ 349  
*P. profundum* DFLVVKSRGEGLEERLAFLEFPDEAVRVEKRPEYDERLEIELKVINQMGPFGYFLIVMEFIQ 348  
*A. hydrophila* DFLVMKSRGEGLEERLEFLFPDPAVRAERPEYDERLDIELKVINQMGPFGYFLIVMEFIQ 349  
*Y. pestis* DFLVEKAKQGLEERLEFLFPDPEVRLQKRPEYDERLDIELKVINQMGPFGYFLIVMEFIQ 360  
*Y. intermedia* DFLVEKSKQGLEERLEFLFPDPEVRVQKRPEYDERLDIELKVINQMGPFGYFLIVMEFIQ 360  
*S. proteamaculans* DFLVLKSKEGLEERLEFLFPDPEVRAQRPEYDERLDIELKVINQMGPFGYFLIVMEFIQ 349  
*E. carotovora* DFLVQCCKKGLERLESLEFPDPEVRVARRPEYDERLDIELQVINQMGPFGYFLIVMEFIQ 349  
*S. glossinidius* DYLVKCKAREGLEERLAFLEFPDADMRREKRPEYDERLAIELQVINQMGPFGYFLIVMEFIQ 349  
*S. enterica Newport* DYLVKKAKEGLEERLAFLEFPDDEERKRRPEYDERLDIELQVINQMGPFGYFLIVMEFIQ 349  
*S. enterica Javiana* DYLVKSKEGLEERLAFLEFPDDEERKRRPEYDERLDIELQVINQMGPFGYFLIVMEFIQ 349  
*S. sonnei* DYLVKRAKEGLEERLAFLEFPDDEERVKRRPEYDERLETELQVINQMGPFGYFLIVMEFIQ 349  
*S. boydii* DYLVKRAKEGLEERLAFLEFPDDEERVKRRPEYDERLETELQVINQMGPFGYFLIVMEFIQ 349  
*E. coli* DYLVKRAKEGLEERLAFLEFPDDEERLKRPEYDERLETELQVINQMGPFGYFLIVMEFIQ 349  
*E. albertii* DYLVKRAKEGLEERLAFLEFPDDEERLKRPEYDVRDLTELQVINQMGPFGYFLIVMEFIQ 349  
*C. koseri* DFLVKKSKEGLEERLAFLEFPDPEVRAKRRPEYDERLDIELQVINQMGPFGYFLIVMEFIQ 333  
*K. pneumoniae* DFLVMKSRGEGLEERLEFLFPDPAVRAEKRRPEYDQRLDIELQVINQMGPFGYFLIVMEFIQ 349  
*E. sakazakii* DFLVMKSKEGLEERLEFLFPDPEIRAQRPEYDERLDIELQVINQMGPFGYFLIVMEFIQ 349  
*P. luminescens* DYLVNRSQQGLEERLAFLEFPDPAVRAEKRRPEYDERLDIELRVINQMGPFGYFLIVMEFIQ 349  
*H. influenzae 3655* DFLVQKSKEGLEERLAFLEFPDPEVRAKRRPEYDERLQVELDVINQMGPFGYFLIVMEFIQ 349  
*H. influenzae 22* DFLVQKSKEGLEERLAFLEFPDPEVRAKRRPEYDERLQVELDVINQMGPFGYFLIVMEFIQ 349  
 \*.\*: :.:\*\*\* \*\* \*:\* \* :.\* \*\* \*\* \* \* \*\*\*\*\*

*V. cholerae* WSKDNDIPVGPGRSGAGSLVAYALKITDLDPLEYDLLFERFLNPERVSMDFDVDFCMD 393  
*V. harveyi* WSKDNAIPVGPGRSGAGSLVAYALKITDLDPLEYDLLFERFLNPERVSMDFDVDFCMD 393  
*V. fischeri* WSKDNAIPVGPGRSGAGSLVAYALDITDLDPLEYDLLFERFLNPERVSMDFDIDFCMD 409  
*P. profundum* WSKDNDVPVGPGRSGAGSLVAYALDITDLDPLKFDLLFERFLNPERVSMDFDVDFCMD 408  
*A. hydrophila* WSKDNGIPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCMD 409  
*Y. pestis* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 420  
*Y. intermedia* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 420  
*S. proteamaculans* WSKDNDVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*E. carotovora* WSKDNDVPVGPGRSGAGSLVAYSALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*S. glossinidius* WSKDNGVVPVGPGRSGAGSLVAYALNITDLDPLAFDLLFERFLNPERVSMDFDVDFCME 409  
*S. enterica Newport* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*S. enterica Javiana* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*S. sonnei* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*S. boydii* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*E. coli* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*E. albertii* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*C. koseri* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 393  
*K. pneumoniae* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*E. sakazakii* WSKDNGIPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*P. luminescens* WSKDNGVVPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCME 409  
*H. influenzae 3655* WSKDNDIPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCMD 409  
*H. influenzae 22* WSKDNDIPVGPGRSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMDFDVDFCMD 409  
 \*\*\*\*\* :\*\*\*\*\* :.\*.\*.\*.\* :\*\*\*\*\* :\*\*\*\*\* :\*\*\*\*\*

*V. cholerae* KRQVDVIDHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 453  
*V. harveyi* KRQVDVIDHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 453  
*V. fischeri* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPSDPGM 469  
*P. profundum* KRDLVIDHVAEYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPSDPGM 468  
*A. hydrophila* RRDEVIDHVSSEMYGREAVSQIITFGSMAAKAVIRDVGRVGLGHAYGFVDRISKLIIPDPGM 469  
*Y. pestis* KRDLVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIPLDPGM 480  
*Y. intermedia* KRDLVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 480  
*S. proteamaculans* KRDLVIDHVAEMYGREAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*E. carotovora* KRDMVIDHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*S. glossinidius* KRDLVIDHVAEYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*S. enterica Newport* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*S. enterica Javiana* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*S. sonnei* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
*S. boydii* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
*E. coli* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
*E. albertii* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
*C. koseri* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 453  
*K. pneumoniae* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*E. sakazakii* KRQDQVIHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLVPPDPGM 469  
*P. luminescens* KRDLVIDHVAEMYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIPLDPGM 469  
*H. influenzae 3655* GRDRVIHVAEYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
*H. influenzae 22* GRDRVIHVAEYGRDAVSQIITFGTMAAKAVIRDVGRVGLGHPYGFVDRISKLIIPDPGM 469  
 \*\* \*.\*.\*.\*: \*\* \*\*\*\*\* :\*\*\*\*\* :\*\*\*\*\* :\*\*\*\*\* :\*\*\*\*\* :\*\*\*\*\*

*V. cholerae* TLEKAFIAEPAALQELYDADEEVKELIDKCRILEGCTRNAGKHAGGVVISPTAITDFAPIY 513  
*V. harveyi* TLEKAFKAEPALPELYDADEEVKELIDMCRLLLEGCTRNAGKHAGGVVISPTTITDFAPIY 513  
*V. fischeri* TLAKAFEAEPQLPELYNADEEVKALIDMCRILEGCTRNAGKHAGGVVISPTTITDFAPLY 529  
*P. profundum* TLAKAFEAEPQLGDAYNADEEVKALIDMCRILEGVTRNAGKHAGGVVISPTTITDFSPLY 528  
*A. hydrophila* TLAKAFEAEPKLPPELYEQDEEVKDLIDMARRLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*Y. pestis* TLEKAFAAEPQLAEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 540  
*Y. intermedia* TLEKAFAAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 540  
*S. proteamaculans* TLEKAFAAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*E. carotovora* TLEKAFAAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*S. glossinidius* TLEKAFVAEPQLQAIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*S. enterica Newport* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*S. enterica Javiana* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*S. sonnei* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*S. boydii* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*E. coli* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*E. albertii* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*C. koseri* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 513  
*K. pneumoniae* TLAKAFEAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*E. sakazakii* TLAKAFEAEPQLQBIYDADEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFAPLY 529  
*P. luminescens* TLEKAFAAEPQLPEIYEADDEEVKALIDMARKLEGVTRNAGKHAGGVVIAPTKITDFSPLY 529  
*H. influenzae 3655* TLSKAFEAEPQLQTAYDSDEEVKALIDMARKLEGVTRNAGKHAGGVVISPTLITDFSPLY 529  
*H. influenzae 22* TLAKAFETEPQLQTAYDSDEEVKALINMARKLEGVTRNAGKHAGGVVISPTLITDFAPLY 529  
 \*\* \*\*\* : \*\* \* \* : \*\*\*\* : \*\* : . \* \*\*\* .\*\*\*\*\* : \*\* \*\*\*\* : \* :

*V. cholerae* CDAEGNFPVVTQFDKNDVETAGLVKFDLGLRSLTIIDWALGLVNPRLKAGKPPVRIEAI 573  
*V. harveyi* ADSEGHFPVVTQFDKNDVETAGLVKFDLGLRSLTIIDWALGLINPRLEREGKDPVRIESI 573  
*V. fischeri* CDADGHPVVTQFDKNDVETAGLVKFDLGLRSLTIIDWALGLINPRLEREGKEPVRIESI 589  
*P. profundum* CDAEGANPVTQFDKNDVETAGLVKFDLGLRSLTIIDWALALINPRLEREGKEPVRINIASI 588  
*A. hydrophila* CDDEGHPVVTQFDKNDVEYAGLVKFDLGLRSLTIIDWALGMINPRLEREGKEPVDIAAI 589  
*Y. pestis* CDAEGNPNVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIASI 600  
*Y. intermedia* CDAEGKNPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIASI 600  
*S. proteamaculans* CDAEGQHPVTQFDKNDVEYAGLVKFDLGLRSLTIIDWALAMINARRAKTGLEPIDIAAI 589  
*E. carotovora* CDSEGNHPVTQFDKNDVEYAGLVKFDLGLRSLTIIDWALEMINARRAKTGLEPIDIATI 589  
*S. glossinidius* CDGEGNHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAQGLEPIDIAAI 589  
*S. enterica Newport* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*S. enterica Javiana* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*S. sonnei* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*S. boydii* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*E. coli* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*E. albertii* CDEEGKHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*C. koseri* CDEAGQHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 573  
*K. pneumoniae* CDEQGLHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*E. sakazakii* CDEMGHHPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*P. luminescens* CDPEGLNPVTQFDKNDVEYAGLVKFDLGLRSLTIINWALEMINARRAKTGLEPIDIAAI 589  
*H. influenzae 3655* CDNEGLHPVTHFDKNDVEYAGLVKFDLGLRSLTIINWALDMINVRMVRGKPRVDIAAI 589  
*H. influenzae 22* CDNEGLHPVTHFDKNDVEYAGLVKFDLGLRSLTIINWALDIINVRMVRGKPRVDIAAI 589  
 . \* \* \*\*\* : \*\*\* : \*\* \* : \*\*\*\*\* : \*\*\* : \* : \* : \* :

*V. cholerae* PLDDARSFRNLQDAKTAVFQLESRGMKELIKRLQPDCEFDI IALVALFRPGPLQSGMVD 633  
*V. harveyi* PLDDAASFRLLQNSETTAVFQLESRGMKELIKRLQPDCEFDI IALVALFRPGPLQSGMVD 633  
*V. fischeri* SLEDQASFRLLQNSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*P. profundum* PMDDKKSFAMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 648  
*A. hydrophila* PIDDKRSFDLLQRSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*Y. pestis* PLEDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 660  
*Y. intermedia* PLEDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 660  
*S. proteamaculans* PLEDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*E. carotovora* PLGDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*S. glossinidius* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*S. enterica Newport* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*S. enterica Javiana* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*S. sonnei* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*S. boydii* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*E. coli* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*E. albertii* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*C. koseri* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 633  
*K. pneumoniae* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*E. sakazakii* PLDDKKSFDMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*P. luminescens* PLNDAKSFAMLQRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*H. influenzae 3655* PLDDPESFELLKRAETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
*H. influenzae 22* PLDDPESFELLKRSSETTAVFQLESRGMKDLIKRLQPDCEFDI IALVALFRPGPLQSGMVD 649  
 . : \* \*\* \* : : \*\*\*\*\* : \*\*\*\*\* : \*\* \* : \*\*\*\*\* : \*\*\*\*\*



*V. cholerae* AIVYGIGAIKGVGEGPIEAILEARNKG--GYFKDLDFDCARIDLKKNVNRVIEKLIILAGA 871  
*V. harveyi* AIVYGIGAIKGVGEGPIDAILEARNKG--GHFIDLDFDCARIDLKKNVNRVIEKLIYAGA 871  
*V. fischeri* AIVYGIGAIKGVGEGPIENII EARNKG--GHFRDLDFDCARIDTKKNVNRVIEKLI RAGA 887  
*P. profundum* AIVYGIGAVKGVGEGPIENLIEAREKG--GHFKDLDFDCARIDTKKNVNRVIEKLI KSGA 886  
*A. hydrophila* HIVYGIGAVKGVGEGPIDAILEARDRD--GPPRDLDFDCNRVDIKKLNKRVMKLIILSGA 887  
*Y. pestis* EIVYGIGAIKGVGEGPIEAILEARKEG--GYFKELFDLFCARVDTKKLNKRILEKLIIMSGA 898  
*Y. intermedia* EIVYGIGAIKGVGEAPEIEAILEARKEG--GHFKDLFDLFCARVDTKKLNKRILEKLIIMSGA 898  
*S. proteamaculans* EIVYGIGAIKGVGEGPIEAILEARNQ--GYFKDLDFDCARSDIKKLNRRVLEKLIIMSGA 889  
*E. carotovora* EIVYGIGAIKGVGEGPIEAII EARNQ--GYFRELFDLFCARTDIKKLNRRVLEKLIIMSGA 887  
*S. glossinidius* EIVYGIGAIKGVGEGPIEAII EARNQ--GHFRELFDLFCARTDIKKLNRRVLEKLIIMSGA 887  
*S. enterica Newport* EIVYGIGAIKGVGEGPIEAII DARNQ--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*S. enterica Javiana* EIVYGIGAIKGVGEGPIEAII EARNQ--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*S. sonnei* EIVYGIGAIKGVGEGPIEAII EARNKG--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*S. boydii* EIVYGIGAIKGVGEGPIEAII EARNKG--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*E. coli* EIVYGIGAIKGVGEGPIEAII EARNKG--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*E. albertii* EIVYGIGAIKGVGEGPIEAII EARNKG--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*C. koseri* EIVYGIGAIKGVGEGPIEAII EARNKD--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 871  
*K. pneumoniae* EIVYGIGAIKGVGEGPIEAII EARNNG--GYFRELFDLFCARTDIKKLNRRVLEKLIIMSGA 887  
*E. sakazakii* EIVYGIGAIKGVGEGPIEAII EARNKD--GYFRELFDLFCARTDTKKLNRRVLEKLIIMSGA 887  
*P. luminescens* EIVYGIGAIKGVGEGPIEAII EARNQ--GYFRELFDLFCARVDTKKLNRRVLEKLIIMSGA 887  
*H. influenzae 3655* EIVYGIGAIKGVGEGPIEALVTARNEG--GIFKDLFDLFCARVLDKKNRRTFESLIIMSGA 887  
*H. influenzae 22* EIVYGIGAIKGVGEGPIEALVAARNEG--GIFKDLFDLFCARVLDKKNRRTFESLIIMSGA 887

\*\*\*\*\*:\*\*\*\*\*:\*. :. :\*. . \* \* :\*\*\*: \* \* \* \* \* : \* . \* . \* \* : \*\*

*V. cholerae* LDRLGPHRAAMMASVDDAVRAASQHHQAEAFQADMPGVLTDAPPEEVEQKYTVQVPEWPEK 931  
*V. harveyi* LDRLGPHRAALMASLNDVAVKAASQHHQAEAFQADMPGVLTDAPPEEVEHKYTVQVPAWPEK 931  
*V. fischeri* LDRLGPHRAALMASLDDAMKAAGQHHQAEAFQADMPGVLTDAPPEEVEQAYIHVPKWPEK 947  
*P. profundum* MDRLGPNRATLMATLNDAIKAAGQHHHAEAFQADMPGVLTDAPPEEVEQAYANIPWPWPEK 946  
*A. hydrophila* MDRLGPHRAALMATLEBAMRAAEQHAQAQAVGVQVDMFGVLTBEIDDKKAFANVPHWPDK 947  
*Y. pestis* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYANVPPWPEQ 958  
*Y. intermedia* FDRLGPHRAALMNSLGEALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYANVPPWPEQ 958  
*S. proteamaculans* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYANVPLWPEQ 949  
*E. carotovora* FDRLGPHRAALMNSLADALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYSTVPPWPEQ 947  
*S. glossinidius* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYSTVPPWPEQ 947  
*S. enterica Newport* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*S. enterica Javiana* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*S. sonnei* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*S. boydii* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*E. coli* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*E. albertii* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPDQ 947  
*C. koseri* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 931  
*K. pneumoniae* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*E. sakazakii* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEQVEQSYASCQPWPEQ 947  
*P. luminescens* FDRLGPHRAALMNSLGDALKAADQHAQAEAIQGVDMFGVLADAPEEVEQSYANISQWPEQ 947  
*H. influenzae 3655* FDKLGPRAALSKNLEDALRASDQHAQAEAIQGVDMFGVLTETHEDVENAYANTPPYTEK 947  
*H. influenzae 22* FDKLGPRAALSKNLEDALRASDQHAQAEAIQGVDMFGVLTETHEDVENAYANTPPYTEK 947

:\*:\*\*\*:\*\*\*: . : :\*:\*: \* \* : : \* \* \*\*\*\*\*: : : : : :

*V. cholerae* VRLEGERETLGLYLTGHPVDEYLKELTKYTS--CRLNEAAPTRRDQSLTVAGLVIAARVMT 990  
*V. harveyi* VWLEGERETLGLYLTGHPVNAYIKELNKVYS--CRLKDATPTRRDQSVTVAGLVIAARVMT 990  
*V. fischeri* VWLEGERETLGLYLTGHPVNAYIKELSRVYTT--WRLNDAHVTRRDQSVTVAGLVIAARVMT 1006  
*P. profundum* VWLEGERETLGLYLTGHPINAYISELKHVYTT--WRLKDNPTGRDKVASVAGLVIAARVMT 1005  
*A. hydrophila* VWLEGERETLGLYLTGHPINQYSGELRRYTS--GRLCDLHPTSRDVTVTAAGLVIAARSMV 1006  
*Y. pestis* VVLDGERETLGLYLTGHPITQYLKEIER YAGGMRLKDMHPTDRGKMTTAVGLVIAARVMV 1018  
*Y. intermedia* IVLDGERETLGLYLTGHPITQYLKEIER YAGGQRLKDMHPTDRGKMTTAVGLVIAARVMV 1018  
*S. proteamaculans* VVLDGERETLGLYLTGHPITQYLKEIER YAGGQRLKDMHPTDRGKMTTAVGLVIAARVMV 1009  
*E. carotovora* VVLDGERETLGLYLTGHPITQYIKIEIER YAAAGVRLKDMHPTDRGKMTTAVGLVIAARVMI 1007  
*S. glossinidius* VVLDGERETLGLYLTGHPITQYLREIER YAGGVRLKDIHPTDRGKMTTAVGLVIAARVMV 1007  
*S. enterica Newport* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*S. enterica Javiana* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*S. sonnei* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*S. boydii* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*E. coli* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*E. albertii* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*C. koseri* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 991  
*K. pneumoniae* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*E. sakazakii* VVLDGERETLGLYLTGHPINQYLKEIER YVGGVRLKDMHPTDRGKMTTAVGLVIAARVMV 1007  
*P. luminescens* IVLDGERETLGLYLTGHPITRYLKEIER YTSGLRLKDVNPTDRGQVTTVIGLVLASKVI I 1007  
*H. influenzae 3655* QILDGERETLGLYLTSSHPVSRYLKELSHYTS--TRKDLAPNRRGQISTVAGLLVASRIAI 1006  
*H. influenzae 22* QILDGERETLGLYLTSSHPVSRYLKELSHYTS--TRKDLAPNRRGQISTVAGLLVASRIAM 1006

\*:\*\*\*\*\*:\*. :. :\*. . \* \* :\*\*\*: \* \* \* \* \* : \* . \* . \* \* : \*\*



## DnaQ SEQUENCES

*V. cholerae* -MNTSNNAEYQRIIVVLDTETTGMNREGGPHYEGHRIIEIGAVEIINRKLTRGRHFHYVYK 59  
*V. harveyi* -MNTSSNSEHNRIVVLDTETTGMNLEGGPHYMGRHRIIEIGAVEIINRKLTRGRHFHYVYK 59  
*V. fischeri* -MNASNNS-SNRIIVLDTETTGMNFSGGPVYEGHRIIVEIGAVEIINRKLTRGRHFHYVYK 58  
*P. profundum* -MNATMKATNQRIIVLDTETTGMNMTG- PHYEGHCIIIEIGAVEIINRKLTRGRHFHYVYK 58  
*A. hydrophila* -MNT---PQLNRQIIVLDTETTGMNTAGGPVYLGHRHRIIEIGAVEIINRKLTRGRHFHYVYK 56  
*Y. pestis* -----MIITPTRQIVLDTETTGMNKLK- VHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*Y. intermedia* MSTGNMQVTPTRQIVLDTETTGMNKLK- VHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 59  
*S. proteamaculans* -----MISTTTRQIVLDTETTGMNKLK- VHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*E. carotovora* -----MSTEITRQIVLDTETTGMNKLK- VHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*S. enterica Newport* --MTDMSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 57  
*S. enterica Javiana* --MTDMSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 57  
*S. boydii* --MTAMSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 57  
*S. sonnei* -----MSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*E. coli* -----MSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*E. albertii* -----MSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*C. koseri* --MTAMSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 57  
*K. pneumoniae* -----MSTAITRQIVLDTETTGMNQIG- AHYEGHKIIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*E. sakazakii* -----MSTAITRQIVLDTETTGMNQIG- AHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*S. glossinidius* -----MSTDITRQIVLDTETTGMNKLK- VHYEGHRIIEIGAVEIINRRLTGRNPHFVYVVK 54  
*P. luminescens* -----MSTAITRQIVLDTETTGMNKLK- VHYEGHKIIIEIGAVEIINRRLTGRHFHYVYIQP 54  
*H. influenzae 22* -----MINPNRQIVLDTETTGMNQIG- AHYEGHCIIIEIGAVEIINRRLTGRNPHFVYVVK 53  
*H. influenzae 3655* -----MINPNRQIVLDTETTGMNQIG- AHYEGHCIIIEIGAVEIINRRLTGRNPHFVYVVK 53  
 \* :\*\*\*\*\*. \* \* \* \* :\*:\*:\*:\*: \* \* :\*:\*:\*

*V. cholerae* DRDIQLEAIEVHGITDEFLDKDPEYKDVHEEFLDFIKGAELVAHNAPPDVGFMDEYEPK 119  
*V. harveyi* DREIQPDAIGVHGITDEFLVDKPEYSAVHQEFLDFIKGAELVAHNAPPDVGFMDEYEPK 119  
*V. fischeri* DRLIDLEAIDVHGITDEFLYDKPEYKDIHDEFLDFIKGAELVAHNAPPDVGFMDEYEPK 118  
*P. profundum* DRTIEVEAIGVHGITDEFLIGKPVYKEVHDEFLEPIKGAELVAHNAPPDVGFMDEYEPK 118  
*A. hydrophila* DRLVDPEAIVHGITDEFLRDKPFSQIADEFIEPIRGAEIAHNAPPDVGFMDEYEPK 116  
*Y. pestis* DRLVDPEAYGVHGISDEFLADKPTFADITPEFLDFIRGAELVIHNAAPDVGFMDEYEP 114  
*Y. intermedia* DRLVDPEAYGVHGISDEFLADKPTFADITPEFLDFIRGAELVIHNAAPDVGFMDEYEP 119  
*S. proteamaculans* DRLVDPEAYGVHGISDEFLADKPTFAQIADEFIDPIRGGELVIHNAAPDVGFMDEYEP 114  
*E. carotovora* DRLVDPEAYNIGHGISDEFLADKPTADVADDFLDFIRGAELVIHNAAPDVGFMDEYEP 114  
*S. enterica Newport* DRLVDPEAFGVHGISDEFLADKPTFADVVDFLDYIRGAELVIHNAAPDVGFMDEYEP 117  
*S. enterica Javiana* DRLVDPEAFGVHGISDEFLADKPTFADVVDFLDYIRGAELVIHNAAPDVGFMDEYEP 117  
*S. boydii* DRLVDPEAFGVHGISDEFLADKPTFAEVADEFMDYIRGAELVIHNAAPDVGFMDEYEP 117  
*S. sonnei* DRLVDPEAFGVHGISDEFLADKPTFAEVADEFMDYIRGAELVIHNAAPDVGFMDEYEP 114  
*E. coli* DRLVDPEAFGVHGISDEFLADKPTFAEVADEFMDYIRGAELVIHNAAPDVGFMDEYEP 114  
*E. albertii* DRLVDPEAFGVHGISDEFLADKPTFADVADEFMDYIRGAELVIHNAAPDVGFMDEYEP 114  
*C. koseri* DRLVDPEAFGVHGISDEFLADKPTFADVVDFLEYIRGAELVIHNAAPDVGFMDEYEP 117  
*K. pneumoniae* DRLVDPEAFGVHGISDEFLADKPTFADVADEFMDYIRGAELVIHNAAPDVGFMDEYEP 114  
*E. sakazakii* DRLVDPEAFMVHGISDEFLMKNKPVFGDVDFLDYIRGAELVIHNAAPDVGFMDEYEP 114  
*S. glossinidius* DRLVDPEAFNVHGISDEFLADKPTFADVADEFHPPIRGGELVIHNAAPDVGFMDEYEP 114  
*P. luminescens* DRLVDPEAFVHGISDEFLQDKPLPADVADEFVPIRGAEIINHNAAPDVGFMDEYEP 114  
*H. influenzae 22* DRSDVDAIKVHGITDEMLADKPEFKETAQDFLDYINGAELLIHNAAPDVGFMDEYEP 113  
*H. influenzae 3655* DRPVDPAIKVHGITDEMLADKPEFKDLAQDFLDYINGAELLIHNAAPDVGFMDEYEP 113  
 \*\* : : \* :\*\*\*:\*:\* \* \* \* : : \* :\*:\*:\*:\*: \* \* \* \* :\*:\*:\* \*

*V. cholerae* SGAIGKTSDFCKVTDTLAMAKRIFPGKRNNLDILCERYGIDNSHRTLHGALLDAEILADV 179  
*V. harveyi* NPSIGKTDYCKVTDTLAMAKKIFPGKRNNLDVLCERYGIDNSHRTLHGALLDAEILADV 179  
*V. fischeri* NKLIGTDDYCKITDTLAMAKKIFPGKRNNLDVLCERYGIDNSHRTLHGALLDAEILADV 178  
*P. profundum* NPNIKGTEDFCKVTDTLAMAKKIFPGKRNNLDILCSRYGIDNSHRTLHGALLDAEILADV 178  
*A. hydrophila* GLNF-KTADICGITDTLAMARDLFPGKRNNLDVLCERYGIDNSHRTLHGALLDAEILADV 175  
*Y. pestis* QQDI PKTETFCITDLSLLMARRLFPGKRNNLDALCDRYQIDNTKRTLHGALLDAEILAEV 174  
*Y. intermedia* QQDI PKTETFCITDLSLLMARRLFPGKRNNLDALCDRYLIDNTKRTLHGALLDAEILAEV 179  
*S. proteamaculans* QQGI PKTDTFCITDLSLLMARRLFPGKRNNLDALCSRYEIDNSKRTLHGALLDAEILAEV 174  
*E. carotovora* NRNI PKTETFCITDLSLPMARKIFPGKRNNLDALCDRYLIDNSKRTLHGALLDAEILAEV 174  
*S. enterica Newport* KRDI PKTNTFCVTDLSLALARKMPFGKRNSLDALCSRYEIDNSKRTLHGALLDAQILAEV 177  
*S. enterica Javiana* KRDI PKTNTFCVTDLSLALARKMPFGKRNSLDALCSRYEIDNSKRTLHGALLDAQILAEV 177  
*S. boydii* KRDI PKTNTFCVTDLSLAVARKMPFGKRNSLDALCARYEIDNSKRTLHGALLDAQILAEV 177  
*S. sonnei* KRDI PKTNTFCVTDLSLAVARKMPFGKRNSLDALCARYEIDNSKRTLHGALLDAQILAEV 174  
*E. coli* KRDI PKTNTFCVTDLSLAVARKMPFGKRNSLDALCARYEIDNSKRTLHGALLDAQILAEV 174  
*E. albertii* KRDI PKTNTFCVTDLSLAVARKMPFGKRNSLDALCARYEIDNSKRTLHGALLDAQILAEV 174  
*C. koseri* SRGI PKTSEFCITDLSLALARRMPFGKRNSLDALCSRYEIDNSKRTLHGALLDAQILADV 177  
*K. pneumoniae* NRGIGKTEFCITDLSLALARKMPFGKRNSLDALCSRYEIDNSKRTLHGALLDAQILADV 174  
*E. sakazakii* NRGIPKTETFCITDLSLALARKMPFGKRNSLDALCSRYEIDNTKRTLHGALLDAQILADV 174  
*S. glossinidius* NRGIAKTDTFCVTDLSLLARKMPFGKRNSLDALCDRYLIDNSKRTLHGALLDAEILADV 174  
*K. pneumoniae* DRDI PPTADFCKITDLSLQARGLFPGKRNNLDALCDRYIDNSKRTLHGALLDAEILADV 172  
*H. influenzae 22* DLNV-KTDDICLVTDTLQMARQMPFGKRNNLDALCDRLGIDNSKRTLHGALLDAEILADV 174  
*H. influenzae 3655* NLNV-KTDDICLVTDTLQMARQMPFGKRNNLDALCDRLGIDNSKRTLHGALLDAEILADV 172  
 . \* \* :\*:\*:\* \* \* \* \* :\*:\*:\*:\*:\*: \* \* \* \* :\*:\*:\*



In particolare, le sequenze aminoacidiche delle 22 subunità polimerasiche sono identiche, in determinate regioni della proteina. Una regione di identità particolarmente interessante sembra essere, per quanto riguarda il prodotto del gene *dnaE* di *Escherichia coli*, quella che va dal residuo V331 a D353 (figura 26). Tale regione comprende la parte terminale della seconda  $\alpha$  elica della regione N-terminale di  $\alpha$  e la  $\alpha$  elica successiva.

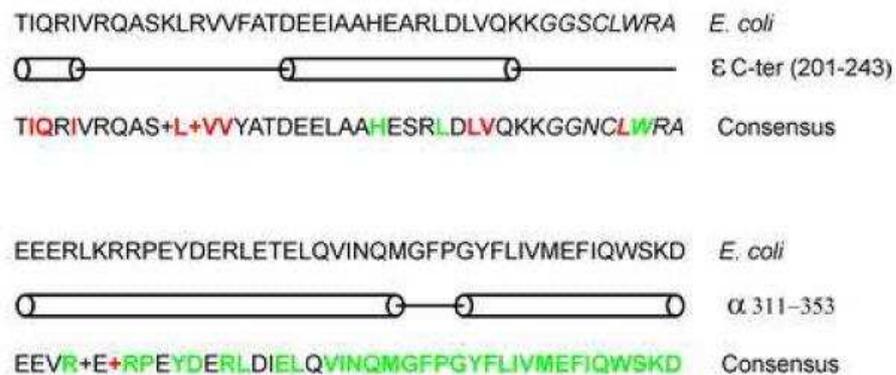


FIG 26 La struttura primaria e secondaria delle regioni delle proteine  $\epsilon$  ed  $\alpha$  che interagiscono per formare il complesso del core della Pol III. L'identità e la conservazione di residui, considerando 22 sequenze provenienti da batteri diversi è indicata rispettivamente in verde e in rosso. I segni + rappresentano un residuo basico. La regione C-terminale di  $\epsilon$  la cui lunghezza varia nei diversi genomi è in carattere italico.

## 8.2 La subunità esonucleasica

Le analisi di allineamento della subunità esonucleasica delle 22 specie batteriche mostrano che solo la regione N-terminale della proteina è altamente conservata, mentre gli aminoacidi che compongono la regione C-terminale sono assolutamente non conservati. Questo probabilmente perché le mutazioni nella regione catalitica di  $\epsilon$  sono nella maggior parte dei casi letali.

Dall' allineamento delle 22 sequenze aminoacidiche della subunità esonucleasica si osserva che la serie di residui che vanno da M189 a A200 in *Escherichia coli* non sono presenti in altre specie batteriche.

Si può ipotizzare che il legame con la subunità polimerasica non avvenga a livello di questa regione.

Nella regione a valle (residui 201- 243) sono invece presenti alcuni residui conservati. In figura 26 si osserva che la regione compresa tra T201 e R213 contiene aminoacidi basici che potrebbero essere coinvolti nel legame con la subunità polimerasica.

### 8.3 Interazione tra $\alpha$ ed $\epsilon$

Gli esperimenti precedentemente descritti dimostrano che la regione C-terminale di  $\epsilon$  è sensibile alla proteolisi e contiene i residui essenziali per l'associazione ad  $\alpha$ . Pertanto l'assemblaggio della Pol III è controllato e dipendente dalla degradazione di  $\epsilon$ . Dal momento che  $\epsilon$ 186 non è in grado di legare  $\alpha$ , mentre  $\epsilon$ 213 può associarsi all'oloenzima, i residui di  $\epsilon$ , compresi tra F187 e R213, sono essenziali per il legame con la subunità polimerasica. L'incapacità di  $\epsilon$ 186 di legarsi in vivo (Taft-Benz e Shaaper, 1999) e in vitro (Perrino et al., 1999) ad  $\alpha$  era già stata evidenziata in letteratura e questo indica che il saggio di attività esonucleasica effettuato utilizzando come substrato il 5'-p-nitrofenil estere della timidina 5'-monofosfato è estremamente valido anche per determinare la attività delle forme di associazione di  $\epsilon$ .

Inoltre è stato dimostrato che  $\epsilon$ 228,  $\epsilon$ 234 e  $\epsilon$ 243 possono legare  $\alpha$  con una minore efficienza rispetto alla variante  $\epsilon$ 213. Anche questo dato evidenzia che i residui responsabili dell'interazione con  $\alpha$  sono compresi tra gli aminoacidi F187 e R213 di  $\epsilon$ .

Inoltre sono state allineate le strutture primarie della subunità  $\epsilon$  appartenente a 22 specie batteriche differenti per le quali è disponibile anche la sequenza di  $\alpha$ . Dall'allineamento delle sequenze aminoacide è possibile notare che la maggiore variabilità si trova immediatamente a valle del dominio esonucleasico di ogni subunità  $\epsilon$ , mentre si osserva un alto grado di conservazione nei residui successivi. Ad esempio, nel caso di *E.coli*, la regione F187-I200 rappresenta la regione variabile, mentre la regione T201-A243 è conservata. Studi recenti (Ozawa et al., 2008) dimostrano che la regione C-terminale di  $\epsilon$ , quando legata ad  $\alpha$ , presenta una regione flessibile (A186-T201) e una regione rigida (I202-A243).

I risultati ottenuti nell'ambito di questa tesi mostrano che la regione compresa tra gli aminoacidi T201 e R213 è essenziale per il legame ad  $\alpha$  e sono pertanto in accordo con i dati disponibili in letteratura. Peraltro, la regione compresa tra T201 e R213 è ricca in aminoacidi basici che potrebbero stabilire delle interazioni elettrostatiche con residui localizzati nella regione di  $\alpha$  che contiene i determinanti per il legame con  $\epsilon$ . Tale regione di  $\alpha$  è stata delimitata (Wieczorek e McHenry, 2006) tra i residui A256 e E320 e, in seguito alle osservazioni riportate in questa tesi, è estesa al residuo D353. Attraverso esperimenti di mutagenesi sito specifica potrebbe essere possibile identificare gli aminoacidi coinvolti nel binding tra  $\epsilon$  ed  $\alpha$ .

## 9. L'assemblaggio della Pol III dipende dalla proteolisi della subunità $\epsilon$

La regione C-terminale di  $\epsilon$ , intrinsecamente instabile, è soggetta a degradazione da parte di proteasi. In figura 27 è rappresentato un modello per l'assemblaggio della DNA polimerasi III in funzione della proteolisi

della subunità  $\epsilon$ . In particolare, lo chaperone molecolare GroL e la proteasi ClpP potrebbero agire insieme per generare forme di  $\epsilon$  intere o parzialmente degradate. Solo alcune varianti degradate di  $\epsilon$  che hanno conservato parte della regione C-terminale (come ad esempio la variante  $\epsilon_{213}$ ) saranno ancora in grado di legare  $\alpha$ . Alternativamente  $\epsilon$  viene protetta da DnaK che preserva  $\epsilon$  dall'azione proteolitica di ClpP e GroL.

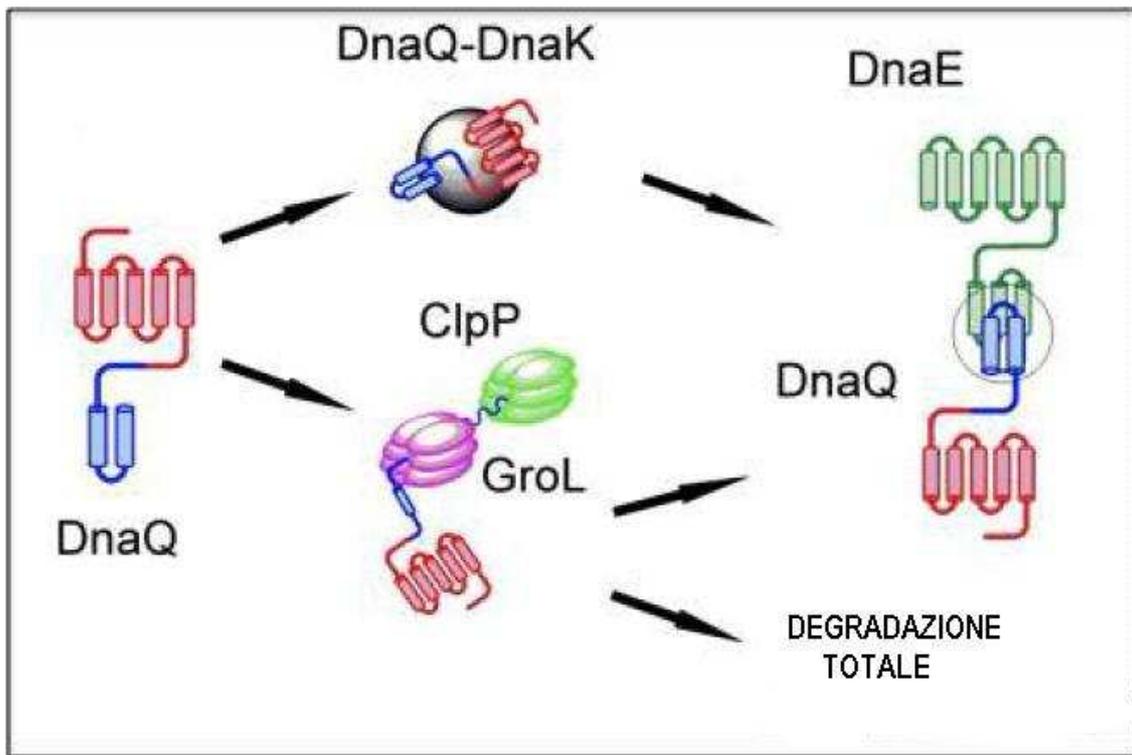


FIG 27: Modello di assemblaggio della DNA Pol III in funzione della stabilità della subunità  $\epsilon$  (regione N-terminale rossa e C-terminale blu). DnaK (grigio), GroL (viola) e ClpP (verde). La regione della subunità  $\alpha$  che interagisce con  $\epsilon$  è indicata con le tre  $\alpha$  eliche verdi.

In letteratura è noto che esiste una forte affinità tra alcune sequenze aminoacidiche e DnaK (Rüdiger et al., 1997). In particolare DnaK si posiziona su una regione di quattro-cinque residui idrofobici fiancheggiati da aminoacidi basici. Invece la presenza di aspartato e di glutammato in una proteina provoca un calo nell'affinità che questa può avere per DnaK probabilmente a causa della natura acida degli aminoacidi presenti nella tasca di legame per il substrato che questo chaperone molecolare presenta (Zhu et al., 1996).

La regione idrofobica V214-A217 della regione C-terminale di  $\epsilon$  è preceduta da residui carichi positivamente e potrebbe essere un sito di legame per DnaK. Questo potenziale sito di legame per DnaK giace subito a valle

del residuo R213 di  $\epsilon$  che fa parte della regione della proteina (T201 e R213) essenziale per il legame con la subunità  $\alpha$  (figura 28).



FIG 28: Sito potenziale di legame tra DnaK (grigio) e la regione C-terminale (residui 201-243) di  $\epsilon$ .

La regione della subunità  $\epsilon$  compresa tra gli aminoacidi V214-A217 potrebbe essere un substrato a cui diverse chaperones molecolari potrebbero legarsi per favorire o ostacolare la degradazione della proteina. Pertanto la subunità  $\epsilon$  potrebbe essere degradata oppure potrebbe formare un complesso con  $\alpha$  prima dell'assemblaggio dell'oloenzima. Altri fattori molecolari, come la subunità  $\theta$ , potrebbero avere un ruolo funzionale nell'assemblaggio della Pol III.

## 10. Il modello di scambio delle polimerasi

Negli ultimi anni sono state individuate nuove DNA Polimerasi sia in eucarioti che in procarioti coinvolte nella replicazione del DNA in condizioni di stress cellulare e indotte dal sistema SOS. In particolare, le DNA polimerasi appartenenti alla famiglia Y (Pol IV e Pol V in *E. coli*) sono state identificate come responsabili della Translesion Synthesis e di elevati livelli di mutagenesi. Poiché queste polimerasi non possiedono una funzione di correzione degli errori e hanno una bassa processività è stato ipotizzato un meccanismo di scambio con le polimerasi ad alta fedeltà definito come "Polymerase Switch Model". In base a tale modello, in *E. coli* le Pol IV e Pol V potrebbero essere temporaneamente reclutate nel momento in cui la sintesi del DNA da parte della DNA Pol III risultasse bloccata per la presenza di lesioni, ad esempio di dimeri di timina, nel DNA stampo. Seguirebbe quindi un breve intervallo di sintesi a bassa fedeltà, dopo il quale verrebbe ripristinata la normale replicazione da parte della DNA Polimerasi III. Lo scambio di tipologia di polimerasi nella sintesi del DNA potrebbe quindi essere essenziale per correggere gli errori, per superare le lesioni presenti nello stampo o per sostituire una polimerasi danneggiata senza dovere disassemblare la struttura del replisoma (Lovett, 2007). Non è stato ancora chiarito se questo processo possa essere un processo coordinato o piuttosto una

competizione tra le polimerasi. Sembra comunque che diverse DNA polimerasi siano impiegate, in maniera specifica, per determinati tipi di lesioni e specifici contesti di sequenza.

Tutte e cinque le polimerasi di *E.coli* legano i  $\beta$  clamps e questa interazione è richiesta perché la sintesi di DNA sia processiva (Lopez de Saro e O'Donnell , 2001). Le DNA Polimerasi potrebbero competere tra loro per il terminale 3' del primer (Napolitano et al., 2000). Lo scambio tra polimerasi potrebbe avvenire in seguito a modificazioni strutturali del DNA, oppure a segnali chimici che si generano da una bolla di replicazione bloccata o, infine, la competizione per lo stesso sito da parte dei  $\beta$  clamps potrebbe innescare il meccanismo di scambio delle polimerasi.

La replicazione da parte di Pol V inoltre è coadiuvata da RecA, a sua volta attivata nella risposta SOS. Al contrario, Pol III viene inibita da RecA. Ci si potrebbe quindi aspettare che Pol III, bloccata in corrispondenza di un sito di lesione del DNA possa essere spiazzata da RecA e allo stesso tempo RecA e la subunità  $\beta$  potrebbero reclutare Pol V. Dopo la riparazione della lesione e il distacco di RecA, dovuto anche alla instabilità di questa proteina, Pol V potrebbe perdere affinità per  $\beta$  e dissociarsi, permettendo la riassociazione di Pol III. I meccanismi competitivi di scambio per l'accesso al DNA non sono ancora stati completamente chiariti, ma bisogna ricordare che le subunità che compongono le polimerasi, come per esempio UmuD e umuC nella Pol V, sono soggette a proteolisi. Tale proteolisi regola finemente il quantitativo cellulare di queste proteine, proprio come, dai dati fino ad ora riportati, avviene per la subunità  $\epsilon$  della DNA Polimerasi III. Pertanto è possibile ipotizzare che esista un meccanismo competitivo regolato a livello della proteolisi delle subunità che compongono le polimerasi. La proteolisi delle singole subunità potrebbe essere il primo passaggio regolatorio della quantità finale delle diverse polimerasi e potrebbe avere le stesse caratteristiche del "Polymerase Switch Model" in modo che la presenza della subunità di una polimerasi escluda la presenza di una componente di un'altra polimerasi.

Il numero di copie di Pol V e di Pol III nella cellula è molto basso soprattutto se rapportato al numero più elevato delle altre polimerasi cellulari. Il numero di copie di Pol III in *E. coli* si aggira intorno a 20 per cellula (Maki et al., 1985) e in condizioni di non induzione della risposta SOS sono circa 15 le molecole di Pol V presenti per cellula (Woodgate and Ennis, 1991).

UmuD e UmuC, subunità che compongono la Pol V sono soggette a proteolisi da parte di Lon e ClpP; la stabilità della subunità  $\epsilon$  della Pol III è contrastata da ClpP e GroL. ClpP potrebbe svolgere un'azione proteolitica aspecifica, mentre le proteine Lon e GroL potrebbero essere gli elementi determinanti che rendono bersaglio-specifica l'azione della proteasi ClpP.

## CONCLUSIONI

Nell'ambito di questa tesi è stata studiata *in vivo* la subunità esonucleasica ( $\epsilon$ ) della DNA polimerasi III (Pol III), enzima preposto alla replicazione del genoma di *Escherichia coli*. La proteina  $\epsilon$  è codificata dal gene *dnaQ* ed è costituita da 243 aminoacidi.

Dagli esperimenti condotti nell'ambito di questa tesi è emerso che il dominio C-terminale di  $\epsilon$  la rende instabile e fortemente soggetta a proteolisi. Sono state realizzate alcune forme troncate di  $\epsilon$  a livello della regione C-terminale e ne è stata studiata la stabilità *in vivo*. In particolare sono state sovraesprese le varianti  $\epsilon$ 186,  $\epsilon$ 200,  $\epsilon$ 213,  $\epsilon$ 228 ed  $\epsilon$ 234.  $\epsilon$ 186 è risultata la variante più stabile ma non in grado di legare la subunità polimerasica.

Sono stati inoltre approfonditi gli aspetti relativi alla stabilità e alla degradazione della proteina  $\epsilon$ . Sovraesprimendo  $\epsilon$  in ceppi di *E. coli* deleti per alcune proteasi o per alcuni chaperones molecolari sono stati identificati i responsabili molecolari della stabilità e della degradazione della subunità esonucleasica. Un ruolo chiave per la emivita di  $\epsilon$  è svolto dalla proteasi ClpP e dagli chaperones molecolari GroL e DnaK. La delezione del gene codificante per la proteina DnaK, per esempio, determina un drammatico decremento di  $\epsilon$ . Inoltre è stata studiata in dettaglio le interazione di  $\epsilon$  con la subunità polimerasica. La subunità polimerasica della Pol III interagisce con la subunità  $\epsilon$  a livello della regione N-terminale e sono state identificate le regioni di  $\alpha$  che legano  $\epsilon$  e che sono coinvolte pertanto nell'assemblaggio dell'oloenzima. La sovraespressione della porzione di  $\alpha$  che interagisce con  $\epsilon$  priva la cellula della subunità enzimatica con attività di correzione delle bozze e produce un fenotipo mutatore. Sono state sovraesprese varianti troncate di  $\alpha$  prive della regione catalitica e, in particolare, le porzioni contenenti i primi 287, 308, 336 e 353 aminoacidi di  $\alpha$ . In tale modo è stato possibile identificare quali sono le regioni di  $\alpha$  che interagiscono con  $\epsilon$ , ossia le  $\alpha$  eliche formate dai residui E311-M335 e G339-D353, indispensabili per il legame ad  $\epsilon$ .

Le informazioni ottenute in questo lavoro di tesi possono essere utilizzate per progettare un potenziale ceppo mutatore condizionale sfruttando l'interazione tra la subunità  $\alpha$  e quella  $\epsilon$ .

Sarebbe, infine, interessante individuare attraverso esperimenti di mutagenesi sito-specifica gli aminoacidi coinvolti nella interazione tra  $\epsilon$  ed  $\alpha$ .

## BIBLIOGRAFIA

Al Mamun A. A. M. and Humayun M. Z. (2008). Spontaneous mutagenesis is elevated in protease-defective cells. *Molecular Microbiology* 71: 629-639.

Aravind L. and Koonin E. V. (1998). Phosphoesterase domains associated with DNA polymerase of diverse origins. *Nucleic Acid Res* 26: 3746-3752.

Arjan J. and De Visser G.M. (2002). The fate of microbial mutators. *Microbiology* 148: 1247-1252.

Baba T., Ara T., Hasegawa M., Takai Y., Okumura Y., Baba M., Datsenk K. A., Tomita M., Wanner B. L. and Mori H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* 1: 1-11

Bachmair A. Finley D. and Varshavsky A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234: 179–186.

Baker T. A. and Bell S. P. (1998). Polymerases and the Replisome: Machines within Machines. *Cell* 92: 295-305.

Baker T. A. and Sauer R. T. (2006). ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* 31:647-653.

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

Battista J.R., Ohta T., Nohmi T., Sun W. and Walker G. C. (1990). Dominant negative *umuD* mutations decreasing RecA-mediated cleavage suggest roles for intact UmuD in modulation of SOS mutagenesis. *Proc Natl Acad Sci USA* 87: 7190-7194.

Bebenek K., Joyce C. M., Fitzgerald M. P. and Kunkel T. A. (1990). The Fidelity of DNA Synthesis Catalyzed by Derivatives of *Escherichia coli* DNA Polymerase I. *J Biol Chem* 265: 13878– 13887.

- Beese L.S., Friedman J.M., Steitz T.A. (1993). Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* 32:14095-14101.
- Berardini M., Foster P.L. and 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.
- Blanco L., Bernard A. and Salas M. (1992). Evidence favoring the hypothesis of a conserved 3'- 5' exonuclease active site in DNA-dependent DNA polymerases. *Gene* 112: 139-144.
- Bonner C.A., Hays S., McEntee K. and 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.
- Bonner C.A., Randall S.K., Rayssiguier C., Radman M., Eritja R., Kaplan B.E., McEntee K. and Goodman M.F. (1988). Purification and characterization of an inducible *Escherichia coli* DNA polymerase capable of insertion and bypass at abasic lesions in DNA. *J Biol Chem* 263 :18946-18952.
- Bonner C.A., Stukenberg P.T., Rajagopalan M., Eritja R., O'Donnell M. and McEntee K. (1992). Processive DNA synthesis by DNA polymerase II mediated by DNA polymerase III accessory proteins. *J Biol Chem* 267: 11431– 11438.
- Bowman G.D., Goedken E. R., Kazmirski S. L., O'Donnell M. and Kuriyan J. (2005) DNA polymerase clamp loaders and DNA recognition. *FEBS Letters* 579: 863–867.
- Bradford M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.
- Bullard J.M., Williams J.C., Acker W.K., Jacobi C., Janjic N., and McHenry C.S. (2002). DNA polymerase III holoenzyme from *Thermus thermophilus* identification, expression, purification of components, and use to reconstitute a processive replicase. *J Biol Chem*. 277:13401-13408.
- Butala M., Zgur-Bertok D. and Busby S. J. W. (2009). The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* 1: 82-93.

- Cai H., McEntee H., Yu K., Kunkel T.A. and 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.
- Camps M. and Loeb L. A.. (2004). When Pol I goes into high gear. Cell Cycle 3: 116-118.
- Chikova A. K. and Schaaper R. M. (2005). The Bacteriophage P1 *hot* Gene Product Can Substitute for the *Escherichia coli* DNA Polymerase III  $\theta$  Subunit. Journal of Bacteriology 187: 5528-5536.
- Courcelle J., Khodursky A., Peter B., Brown P.O. and Hanawalt P.C. (2001). Comparative gene expression profiles following UV exposure in *wild-type* and SOS-deficient *Escherichia coli*. Genetics 158: 41–64.
- Cox E.C. and Horner D.L. (1986). DNA sequence and coding properties of *mutD* (*dnaQ*) a dominant *Escherichia coli* mutator gene. J Biol Chem 190: 113-117.
- Dallmann H.G., Thimmig R.L. and 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.
- Dame R. T. and Goosen N. (2002). HU: promoting or counteracting DNA compaction? FEBS Letters 529: 151-156.
- De Rose E. F., Darden T., Harvey S., Gabel S., Perrino F. W., Schaaper R. M and London R. E. (2003). Elucidation of the  $\epsilon$ - $\theta$  Subunit Interface of *Escherichia coli* DNA Polymerase III by NMR Spectroscopy. Biochemistry 42: 3635-3644.
- Deuerling E., Schulze-Specking A., Tomoyasu T., Mogk A. and Bukau B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400, 693-696.
- Donnelly C.E. and Walker G.C. (1992). Coexpression of UmuD' with UmuC suppresses the UV mutagenesis deficiency of *groE* mutants. J. Bacteriol. 174 3133–3139.
- Dorson J. W., Deutsch W. A. and Mose R. E. (1978). Role of DNA Polymerases in Excision Repair in *Escherichia coli*. J. Biol.Chem 253 : 660-664.

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

Dri A. M. and Moreau P.L.. (1994). Control of the LexA regulon by pH: evidence for a reversible inactivation of the LexA repressor during the growth cycle of *Escherichia coli*. *Mol Microbiol* 12: 621–629.

Ellis R. J. (1993). The general concept of molecular chaperones. *Phil Trans R Soc Lond* 339: 257-261.

Erbse A., Schmidt R., Bornemann T., Schneider-Mergener J., Mogk A., Zahn R., Dougan D.A. and Bukau B. (2006). ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. *Nature* 439: 753–756.

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

Farr G.W., Furtak K., Rowland M. B., Ranson N. A., Saibil H. R., Kirchhausen T. and Horwich A. L. (2000). Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* 100: 561- 573.

Flower A. M. and McHenry C. S. (1990). The  $\gamma$  subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frame shifting. *Proc Natl Acad Sci USA* 87:3713-3717.

Foster P.L. and Eisenstadt E. (1985). Induction of transversion mutations in *Escherichia coli* by N-methyl-N'-nitro-N-nitrosoguanidine is SOS dependent. *J Bacteriol* 163: 213-220.

Foster P. L. and Marinus M. G. (1992). Levels of  $\epsilon$ , an essential replication subunit of *Escherichia coli* DNA polymerase III, are controlled by heat shock proteins. *J. Bacteriol* 174, 7509-7516.

Foster P. L. (2004). Stress responses and genetic variation in bacteria. *Mutation Research* 569: 3–11.

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

Frank E. G., Gonzalez M., Ennis D. G, Levine A. S. and Woodgate R. (1996). In Vivo Stability of the Umu Mutagenesis Proteins: a Major Role for RecA. *Journal of Bacteriology* 178: 3550–3556.

Fritsch E.F., Sambrook J. and Maniatis T. (1989). *Molecular Cloning*. Cold Spring Harbor.

Fuchs R. P., Fujii S. and Wagner J. (2004). Properties and Functions of *Escherichia coli*: Pol IV and Pol V. *Adv Protein Chem.* 69: 229-264.

Gao D. and McHenry C. S. (2001).  $\tau$  Binds and Organizes *Escherichia coli* Replication Proteins through Distinct Domains. *J. Biol.Chem* 276: 4433–4440.

Gao D. and McHenry C. S. (2001).  $\tau$  Binds and Organizes *Escherichia coli* Replication Proteins through Distinct Domains. *J. Biol.Chem* 276: 4441–4446.

Giglione C., Boularot A. and Meinnel T. (2004). Protein N-terminal methionine excision. *Cell Mol Life Sci* 61: 1455–1474.

Glover B. P. and McHenry C. S. (1998). The  $\chi\psi$  Subunits of DNA Polymerase III Holoenzyme Bind to Single-stranded DNA-binding Protein (SSB) and Facilitate Replication of an SSB-coated Template. *J. Biol.Chem* 273: 23476–23484.

Glover B. P. and McHenry C. S. (2001). The DNA Polymerase III Holoenzyme: An Asymmetric Dimeric Replicative Complex with Leading and Lagging Strand Polymerases. *Cell* 105: 925–934.

Glover B. P., Pritchard A. E. and McHenry C. S. (2001).  $\tau$  Binds and Organizes *Escherichia coli* Replication Proteins through Distinct Domains. *J. Biol.Chem* 276: 35842–35846.

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

Gonzalez M., Frank E. G., Levine A. S. and 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.

Gonzalez M., Resulova F., Maurizi M. R. and Woodgate R. (2000). Subunit-specific degradation of the UmuD/D' heterodimer by the ClpXP protease: the role of *trans* recognition in UmuD' stability. *EMBO Journal* 19: 5251-5258.

Gottesman S. (1996). Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* 30: 465–506.

Gottesman S., Roche E., Zhou Y. and Sauer R.T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 12: 1338-1347.

Gottesman S. (2003). Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol.* 19: 565-587.

Hamdan S. (2002). Hydrolysis of the 5'-p-nitrophenil Ester of TMP by the proofreading exonuclease ( $\epsilon$ ) subunit of *Escherichia coli* DNA polymerase III. *Biochemistry*, 41: 5266-5275.

Hengge-Aronis R. (2002). Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* 66: 373–395.

Hersh M. N., Ponder R. G., Hastings P.J. and Rosenberg S. M. (2004). Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. *Research in Microbiology* 155: 352–359

Hisaji M., Takashi H. and Mutsuo S. (1983). Structure and expression of the *dnaQ* mutator and *Rnase H* genes of *Escherichia coli*: overlap of the promoter regions. *Proc Natl Acad Sci USA* 80: 7137-7141.

Horwich A.L., Farr G.W. and Fenton W.A. (2006). GroEL-GroES-mediated protein folding. *Chem Rev* 106: 1917-1930.

Iyer R.R., Pluciennik A., Rosche W.A., Sinden R.R. and Wells R.D. (2000). DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in *Escherichia coli*. *J Biol Chem* 275: 2174-2184.

Janion C. (2008). Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. *J Biol Sci* 4: 338-344.

Jarosz D. F., Beuning P.J., Cohen S. E. and Walker G.C. (2007). Y-family DNA polymerases in *Escherichia coli*. *TRENDS in Microbiology* 15: 70-77.

- Jeruzalmi D., O'Donnell M. and Kuriyan J. (2001). Crystal Structure of the Processivity Clamp Loader Gamma ( $\gamma$ ) Complex of *E. coli* DNA Polymerase III. *Cell* 106: 429–441.
- Johanson K. O. and McHenry C. S. (1984). Adenosine 5'-0-(3-Thiotriphosphate) Can Support the Formation of an Initiation Complex between the DNA Polymerase III Holoenzyme and Primed DNA. *J Biol Chem* 259: 4589-4595.
- Johnson A. and O'Donnell M. (2005). Cellular DNA replicases: Components and Dynamics at the Replication Fork. *Annu Rev Biochem* 74: 283–315.
- 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.
- Jones D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195- 202.
- Joyce C. M, Steitz T. A. (1994). Function and structure relationships in DNA polymerases. *Annu Rev Biochem* 63: 777-822.
- Kelman Z., Yuzhakov A., Andjelkovic J. and O'Donnell M. (1998). Devoted to the lagging strand-the subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J.* 17: 2436-2449.
- Keniry M. A., Park, A. Y., Owen E. A., Hamdan S. M., Pintacuda G., Otting G. and Dixon N. E. (2006). Structure of the  $\theta$  Subunit of *Escherichia coli* DNA Polymerase III in Complex with the  $\varepsilon$  Subunit. *Journal of Bacteriology* 188: 4464–4473.
- Kenniston J.A., Baker T.A., Fernandez J.M. and Sauer R.T. (2003). Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* 114: 511-520.
- Kim S., Dallmann H. G, McHenry C. S. and Marians K. J. (1996). Coupling of a Replicative Polymerase and Helicase: A  $\tau$ -DnaB Interaction Mediates Rapid Replication Fork Movement. *Cell* 84: 643–650.

- Kim S.R., Matsui K., Yamada M., Gruz P. and 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.
- Knippers R. (1970). DNA polymerase II. *Nature (London)* 228: 1050–1053.
- Kornberg A. (2000). Ten Commandments: Lessons from the Enzymology of DNA Replication. *Journal of Bacteriology* 182: 3613–3618.
- Kornberg A. and Baker T.A. (2005). DNA replication (second edition) University Science Books.
- Kunkel T. and Bebenek K. (2000). DNA Replication Fidelity. *Annu Rev Biochem* 69: 497–529.
- Kuriyan J. and O'Donnell M. (1993). Sliding clamps of DNA polymerases. *J Mol Biol* 234: 915-925.
- Lamers M. H., Georgescu R. E., Lee S-G., O'Donnell M. and Kuriyan J. (2006). Crystal Structure of the Catalytic  $\alpha$  Subunit of *E. coli* Replicative DNA Polymerase III. *Cell* 126: 881-892.
- Langston L. D. and O'Donnell M. (2006). DNA Replication: Keep Moving and Don't Mind the Gap. *Molecular Cell* 23: 155–160.
- Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez, R., Thompson J. D., Gibson T. J. and Higgins D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- Layton J.C. and Foster P.L. (2003). Error-prone DNA polymerase IV is controlled by the stress-response sigma factor, RpoS, in *Escherichia coli*. *Mol Microbiol* 50: 549–561.
- Layton J.C and Foster PL. (2005). Error-prone DNA polymerase IV is regulated by the heat shock chaperone GroE in *Escherichia coli*. *J Bacteriol* 187: 449-457.
- LeClerc J. E., Payne W. L., Kupchella E. and Cebula T. A. (1998). Detection of mutator subpopulations in *Salmonella typhimurium* LT2 by reversion of his alleles. *Mutat Res* 400: 89-97.
- Lee J., Chastain P.D., Kusakabe T., Griffith J.D. and Richardson C.C. (1998). Coordinated leading and lagging

strand DNA synthesis on a minicircular template. *Mol Cell* 1: 1001-1110.

Lee I., Berdis A.J. and Suzuki C.K. (2006). Recent developments in the mechanistic enzymology of the ATP-dependent Lon protease from *Escherichia coli*: highlights from kinetic studies. *Mol Biosyst* 2:477-483.

Lehman I. R., Bessman M. J., Simms E. S. and Kornberg A. (1958). Enzymatic Synthesis of Deoxyribonucleic Acid. *J.Biol.Chem* 233: 163-170.

Leonard A. C. and Grimwade J. E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Molecular Microbiology* 55: 978–985.

Leu F. P., Georgescu R. and O'Donnell M. (2003). Mechanism of the *E. coli*  $\tau$  Processivity Switch during Lagging-Strand Synthesis. *Molecular Cell* 11: 315–327.

Li X. and Marians K.J. (2000). Two distinct triggers for cycling of the lagging strand polymerase at the replication fork. *J Biol Chem* 275: 34757-34765.

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.

Licht S. and Lee I. (2008) Resolving Individual Steps in the Operation of ATP-Dependent Proteolytic Molecular Machines: From Conformational Changes to Substrate Translocation and Processivity. *American Chemical Society* 47: 3595- 3605.

Little J.W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* 73: 411–421.

López de Saro F. J. and O'Donnell M. (2001). Interaction of the  $\beta$  sliding clamp with MutS, ligase, and DNA polymerase I. *Proc Natl Acad Sci USA* 98: 8376–8380.

Lovett S.T. (2007). Polymerase switching in DNA replication. *Mol Cell* 25: 523-6.

Lupas A.N. and Koretke K.K. (2003). Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. *J Struct Biol* 141: 77–83.

- Maki H., Horiuchi T. and Kornberg A. (1985). The polymerase subunit of DNA polymerase III of *Escherichia coli*. I. Amplification of the *dnaE* gene product and polymerase activity of the alpha subunit. J Biol Chem 260: 12982-12986.
- Maki H. and Kornberg A. (1985). The polymerase subunit of DNA polymerase III of *Escherichia coli*. II. Purification of the alpha subunit, devoid of nuclease activities. J Biol Chem 260: 12987-12992.
- Maki H. and 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, Maki S. and Kornber A. (1988). DNA Polymerase III Holoenzyme of *Escherichia coli*. J Biol Chem 263: 6570-6578.
- Mao E. F., Lane L., Lee J. and Miller J. H. (1997). Proliferation of mutators in a cell population. J Bacteriol 179: 417-422.
- Masker W., Hanawalt P. and Shizuya H. (1973). Role of DNA polymerase II in repair replication in *Escherichia coli*. Nat New Biol 244: 242-243.
- McGinness K.E., Baker T.A. and Sauer R.T. (2006). Engineering controllable protein degradation. Mol Cell 22: 701-707.
- McHenry C.S. (1985). DNA polymerase III holoenzyme of *Escherichia coli*: components and function of a true replicative complex. Mol Cell Biochem 66: 71-85.
- McInerney P. and O'Donnell M. (2004). Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. J Biol Chem 279: 21543-21551.
- Miller J. H. (1992). A short course in bacterial genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.
- Miller J.H. (1998). Mutators in *Escherichia coli*. Mutat. Res. 409: 99-106.
- Miller J. H., Anjali Suthar T. J., Yeung A., Truong C. and Stewart J. L. (1999). Direct selection for mutators in *Escherichia coli*. Journal of Bacteriology 181: 1576-1584.

- Minnick D.T., Astatke M., Joyce C. M. and 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.
- Mogk A., Schmidt R. and Bukau B. (2007). The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol* 17: 165-172.
- Morita M.T., Kanemori M., Yanagi H. and Yura T. (2000). Dynamic interplay between antagonistic pathways controlling the sigma 32 level in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 97:5860-5865.
- Moser M.J., Holley W.R., Chatterjee A. and Mian I.S. (1997). The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucl Acids Res* 25: 5110-5118.
- Naktinis V., Onrust R., Fang L. and 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.
- Napolitano R., Janel-Bintz R., Wagner J. and Fuchs R. P. P. (2000). All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *The EMBO Journal* 19: 6259-6265.
- Neher S. B., Sauer R.T. and Baker T.A.(2003). Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease. *Proc Natl Acad Sci USA* 100: 13219–13224.
- Neylon C., Kralicek A. V., Hill T. M. and Dixon N. E. (2005) Replication Termination in *Escherichia coli*: Structure and Antihelicase Activity of the Tus-Ter Complex. *Microbiology And Molecular Biology Reviews* 69: 501-526.
- Nishii W., Maruyama T., Matsuoka R., Muramatsu T. and Takahashi K. (2002). The unique sites in Sula protein preferentially cleaved by ATP-dependent Lon protease from *Escherichia coli*. *Eur J Biochem.* 269: 451-457.
- O'Donnell M. (2006). Replisome architecture and dynamics in *Escherichia coli*. *J Biol Chem* 281: 10653–10656.

- Ohmori H., Friedberg E.C., Fuchs R. P.P., Goodman M. F., Hanaoka F., Hinkle D., Kunkel T. A., Lawrence C. W., Livneh Z., Nohmi T., Prakash L., Prakash S., Todo T., Walker G. C., Wang Z. and Woodgate R. (2001). The Y-family of DNA Polymerases. *Molecular Cell* 8: 7-8.
- Ohmori H., Hatada E., Qiao Y., Tsuji M. and Fukuda R. (1995). *dinP*, a new gene in *Escherichia coli*, whose product shows similarities to UmuC and its homologues. *Mutat Res.* 347: 1-7.
- Okazaki R., Arisawa M., and Sugino A. (1971). Slow joining of newly replicated DNA chains in DNA polymerase I-deficient *Escherichia coli* mutants. *Proc Natl Acad Sci USA* 68: 2954–2957.
- Ozawa K., Jergic S., Park A.Y., Dixon N.E. and Otting G. (2008). The proofreading exonuclease subunit epsilon of *Escherichia coli* DNA polymerase III is tethered to the polymerase subunit alpha via a flexible linker. *Nucleic Acids Res* 36: 5074-5082.
- Paek K. H. and Walker G. C. (1987). *Escherichia coli dnaK* null mutants are inviable at high temperature. *J Bacteriol* 169: 283-290.
- Patel H. P. and Loeb L. A. (2001). Getting a grip on how DNA polymerases function. *Nat Struct Biol* 8: 656-659.
- Patel H. P., Suzuchi M., Adman E., Shinkai A. and Loeb L. A. (2001). Prokaryotic DNA Polymerase I: Evolution, Structure, and "Base Flipping" Mechanism for Nucleotide Selection. *J Mol Biol* 308: 823-837.
- Perrino F., Harvey S. and McNeill M. (1999). Two functional domains of the  $\epsilon$ -subunit of DNA polymerase III. *Biochemistry* 38: 16001-16009.
- Pham P., Bertram J. G., O'Donnell M., Woodgate R. and Goodman M. F. (2001). A model for SOS-lesion-targeted mutations in *Escherichia coli*. *Nature* 414: 366-370.
- Pham P.T., Olson M.W., McHenry C.S. and Schaaper R.M. (1999). Mismatch extension by *Escherichia coli* DNA polymerase III holoenzyme. *J Biol Chem* 274: 3705-3710.
- Qiu Z. and Goodman M. F. (1997). The *Escherichia coli polB* Locus Is Identical to *dinA*, the Structural Gene for DNA Polymerase II. *J Biol Chem* 272: 8611–8617.

- Radman M. (1999). Enzymes of evolutionary change. *Nature* 401: 866-869.
- Rangarajan S., Woodgate R. and Goodman M. F. (1999). A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated *Escherichia coli*. *Proc Nat Acad Sci U.S.A.* 96: 9224–9229.
- Rüdiger S., Germeroth L., Schneider-Mergener J. and Bukau B. (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* 16, 1501-1507.
- Saibil H. R. (2008). Chaperone machines in action *Current Opinion in Structural Biology* 18: 35–42.
- Salinas F. and Benkovic S.J. (2000). Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate. *Proc Natl Acad Sci U S A.* 97: 7196-7201.
- Sauer R.T., Bolon D.N., Burton B.M., Burton R.E., Flynn J.M., Grant R.A., Hersch G.L., Joshi S.A., Kenniston J.A., Levchenko I., Neher S.B., Oakes E.S., Siddiqui S.M., Wah D.A. and Baker T.A. (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell.* 119: 9-18.
- Schaaper R.M. (1993). Base selection, proofreading and mismatch repair during DNA replication in *E. coli*, *J Biol Chem* 268:23762-23765.
- Schaaper R.M. (1998). Antimutator Mutants in Bacteriophage T4 and *Escherichia coli* *Genetics* 148: 1579–1585.
- Schlacher K., Leslie K., Wyman C., Woodgate R., Cox M. M. and Goodman M.F. (2005). DNA Polymerase V and RecA Protein, a Minimal Mutasome. *Molecular Cell* 17: 561–572.
- Selifonova O., Valle F. and Schellenberger V. (2001). Rapid evolution of novel traits in microorganisms. *Appl Environ Microbiol.* 67: 3645-3649.
- Setlow P., Brutlag D. and Kornberg A. (1972). Deoxyribonucleic Acid Polymerase: two distinct enzymes in one polypeptide. *J Biol Chem* 247: 224-231.

- Slater S. C., Lifshits M. R., O'Donnell M. and Russell M. (1994). *hoIE*, the Gene Coding for the  $\theta$  Subunit of DNA Polymerase III of *Escherichia coli*: Characterization of a *hoIE* Mutant and Comparison with a *dnaQ* ( $\epsilon$ -Subunit) Mutant. *Journal of Bacteriology* 176: 815-821.
- Smith C.K., Wöhnert J., Sauer R.T. and Schwalbe H. (2001). Assignments of the  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  resonances of the substrate-binding SSD domain from Lon protease. *J Bio mol NMR* 21:387-388.
- Stano N. M., Chen J., McHenry C. S. (2006). A coproofreading  $\text{Zn}(2+)$ -dependent exonuclease within a bacterial replicase. *Nat Struct Mol Biol* 13: 458-459.
- Stefan A., Reggiani L., Cianchetta S., Radeghieri A., Gonzalez Vara y Rodriguez A. and Hochkoepler A. (2003). Silencing of the gene coding for the epsilon subunit of DNA polymerase III slows down the growth rate of *Escherichia coli* populations. *FEBS Lett.* 546: 295-299.
- Studwell-Vaughan P. S. and O'Donnell M. (1993). DNA Polymerase III Accessory Proteins. *J Biol Chem* 268: 11785-11791.
- Taddei F., Matic I., Radman M. (1995). cAMP-dependent SOS induction and mutagenesis in resting bacterial populations. *Proc Natl Acad Sci USA.* 92: 11736–11740.
- Taft-Benz S.A. and Schaaper R.M. (1998). Mutational analysis of the  $3' \rightarrow 5'$  proofreading exonuclease of *Escherichia coli* DNA polymerase III. *Nucleic Acids Res.* 26: 4005-4011.
- Taft-Benz S.A. and Schaaper R.M. (1999). The C-Terminal domain of DnaQ contains the polymerase binding site. *J Bacteriol* 181: 2963-2965.
- Taft-Benz S.A. and Schaaper R.M (2004). The  $\theta$  Subunit of *Escherichia coli* DNA Polymerase III: a Role in Stabilizing the  $\epsilon$  Proofreading Subunit. *J Biol Chem* 186: 2774–2780.
- Taguchi H. (2005) Chaperonin GroEL meets the substrate protein as a "load" of the rings. *J Biochem* 137: 543-549.
- Tang M., Pham P., Shen X., Taylor J.-S., O'Donnell M., Woodgate R. and Goodman M. (2000) Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* 404: 1014–1018.

Tang M., Shen X., Frank E.G., O'Donnell M., Woodgate R. and Goodman M.F. (1999) UmuD'<sub>2</sub>C is an error-prone DNA polymerase, *Escherichia coli*, DNA pol V. Proc Natl Acad Sci USA 96: 8919–8924.

Theobald D. L, Mitton-Fry R. M. and Wuttke D.S. (2003). Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct 32:115-133.

Tobias J.W., Shrader T.E., Rocap G. and Varshavsky A. (1991). The N-end rule in bacteria. Science 254, 1374– 1377.

Tompkins J.D., Nelson J.L., Hazel J.C., Leugers S.L., Stumpf J.D. and Foster P.L. (2003) Error-prone polymerase, DNA polymerase IV, is responsible for transient hypermutation during adaptive mutation in *Escherichia coli*. J Bacteriol 185: 3469–3472.

Tougu K., Marians K.J. (1996). The extreme C- terminus of primase is required for interaction with DnaB at the replication fork. J Biol Chem 271: 21391-21397.

Tougu K. and Marians K.J. (1996). The interaction between helicase and primase sets the replication fork clock. J Biol Chem 271: 21398-21405.

Viitanen P.V., Gatenby A.A. and Lorimer G.H.(1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. Protein Sci 1:363-369.

Wagner J., Gruz P., Kim S.R., Yamada M., Matsui K., Fuchs R.P.P. and Nohmi T. (1999). The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA Pol IV, involved in mutagenesis. Mol Cell 4: 281–286.

Wang T.C. (2005). Discontinuous or semi-discontinuous DNA replication in *Escherichia coli*? Bioessays 27: 633–636.

Wieczorek A. and McHenry C. S. (2006). The NH<sub>2</sub>-terminal php Domain of the {alpha} Subunit of the *Escherichia coli* Replicase Binds the epsilon Proofreading Subunit. J Biol Chem 281:12561-12567.

Woodgate R. (1992). Construction of a *umuDC* operon substitution mutation in *Escherichia coli*. Mutat Res 281: 221–225.

- Woodgate R., and Ennis, D.G. (1991). Levels of chromosomally encoded Umu proteins and requirements for *in vivo* UmuD cleavage. *Mol Gen Genet* 229: 10–16.
- Wootton J.C. and Drummond M.H. (1989). The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng* 2: 535-543.
- Xu Y., Derbyshire V., Ng K, Chen Sun X., Grindley N.D. F. and Joyce C. M. (1997). Biochemical and Mutational Studies of the 5'→3' Exonuclease of DNA Polymerase I of *Escherichia coli*. *J Mol Biol* 268: 284-302.
- Yang J., Nelson S.W. and Benkovic S.J. (2006). The control mechanism for lagging strand polymerase recycling during bacteriophage T4 DNA replication. *Mol Cell* 21: 153–164.
- Yao N.Y. and O'Donnell M. (2008) Replisome dynamics and use of DNA trombone loops to bypass replication blocks. *Mol Biosyst* 4: 1075-1084.
- Yeiser B., Pepper E.D., Goodman M.F. and Finkel S.E. (2002). SOS-induced DNA polymerases enhance longterm survival and evolutionary fitness. *Proc Natl Acad Sci USA* 99: 8737–8741.
- Yuzhakov A., Kelman Z. and 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.
- Zhu X., Zhao X., Burkholder W. F., Gragerov A., Ogata C. M., Gottesman M. E. and Hendrickson W. A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 272: 1606-1614.