UNIVERSITA' DEGLI STUDI DI NAPOLI
"FEDERICO II"

DOTTORATO DI RICERCA IN GENETICA E MICROBIOLOGIA CELLULARE E MOLECOLARE

TESI

"SEQUENZE INTERGENICHE E CONTROLLO POST-TRASCRIZIONALE IN MICRORGANISMI PATOGENI"

Coordinatore
Prof. Carmelo Bruno Bruni

Candidato
Dott. Eliana De Gregorio

ANNO
2004

UNIVERSITA' DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI BIOLOGIA E PATOLOGIA CELLULARE E MOLECOLARE "L. CALIFANO"

TESI DI DOTTORATO IN GENETICA E MICROBIOLOGIA CELLULARE E MOLECOLARE XVI CICLO

"SEQUENZE INTERGENICHE E CONTROLLO POST-TRASCRIZIONALE IN MICRORGANISMI PATOGENI"

CANDIDATO: Dott. Eliana De Gregorio

DOCENTE GUIDA: Prof. PierPaolo Di Nocera

UNIVERSITA' DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI BIOLOGIA E PATOLOGIA CELLULARE E MOLECOLARE "L. CALIFANO"

DOTTORATO IN GENETICA E MICROBIOLOGIA CELLULARE E MOLECOLARE

COORDINATORE DEL CORSO DI DOTTORATO: PROF. CARMELO BRUNO BRUNI

Sede Amministrativa:
Università degli Studi di Napoli "Federico II"

Dipartimenti concorrenti:
Biochimica e Biotecnologie Mediche

Curricula del Dottorato:

1. **Genetica e Microbiologia Cellulare**
2. **Genetica e Microbiologia Molecolare**

Collegio dei Docenti

Prof. Carmelo Bruno Bruni: Coordinatore del dottorato
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano", Università di Napoli

Prof. Stefano Bonatti
Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli

Prof. Maria Stella Carlomagno
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano" Università di Napoli

Prof. Sergio Cocozza
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano" Università di Napoli

Prof. Roberto Di Lauro
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano" Università di Napoli

Prof. Paola Di Natale
Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli

Prof. Pier Paolo Di Nocera
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano" Università di Napoli

Prof. Tommaso Russo
Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli

Dott. Nicola Zambrano
Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli

Dott. Raffaele Zarrilli
Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano" Università di Napoli

INDICE

INTRODUZIONE

La rapidità con cui sono state determinate le sequenze complete di oltre 150 genomi batterici e la possibilità di effettuare in breve tempo una varietà di analisi bioinformatiche permette di studiare in dettaglio la distribuzione e le proprietà di famiglie di DNA ripetuto presenti nei genomi procariotici.

La maggior parte del genoma degli organismi procariotici (85-90%) è costituita da sequenze codificanti. Una significativa frazione del DNA intergenico che separa le ORFs (open reading frames) è costituito da sequenze di DNA ripetuto.

La principale classe di sequenze ripetute presente nel genoma di organismi procariotici è rappresentata dalle sequenze IS (Insertion Sequences), elementi genetici mobili caratterizzati da lunghe sequenze terminali ripetute invertite (TIRs) di 10-40 coppie di basi (bp). Le sequenze di inserzione variano in dimensione da 750 a 2500 bp e codificano per una trasposasi che è necessaria per la loro trasposizione. L'integrazione delle IS determina spesso una duplicazione del sito bersaglio; di conseguenza, analisi dei siti di inserzione mostrano che una IS è sempre fiancheggiata da brevi ripetizioni dirette (DR), la cui lunghezza è caratteristica per un dato elemento e varia da 2 a 14 bp. Le sequenze IS sono capaci di promuovere vari tipi di riarrangiamenti genetici, quali delezioni e inversioni, che portano

all'assemblaggio di clusters di geni con funzioni specializzate. In aggiunta le IS possono essere coinvolte nell'attivazione o nell'abolizione dell'espressione genica (Mahillon et al., 1999).

I genomi dei procarioti contengono differenti classi di sequenze ripetute intergeniche non codificanti di dimensioni (40-300 bp) notevolmente inferiori a quelle delle IS (Bachellier et al., 1999).

Le sequenze note come *REP* (Repetitive Extragenic Palindromic sequences, Stern et al., 1984) o *PU* (Palindromic Units, Higgins et al., 1982), sono palindromi imperfette di circa 33-40 bp (Higgins et al., 1982). Nei genomi di *Escherichia coli* e *Salmonella typhimurium,* il numero di elementi *PU* oscilla tra le 500 e le 1000 copie, contribuendo a formare una significativa frazione del genoma (0.5-1%). In numero minore, sequenze *PU* sono presenti in *Klebsiella pneumoniae* ed altri enterobatteri. Gli elementi *PU* variano in sequenza in maniera specie-specifica, conservando la capacità di formare strutture stem-loop stabili a livello di RNA. Le sequenze *PU* si ritrovano in singola copia o ripetuti in tandem in orientamento invertito. Esistono due tipi di elementi *PU*, chiamati *Y* e *Z*, differenti fra loro per il nucleotide presente nella settima e trentaduesima posizione (Figura 1A). Le sequenze di tipo *Z* sono ulteriormente suddivisibili in due tipi in base alla lunghezza (Figura 1A). Delle 500 sequenze *PU* presenti

in *E. coli* solo 83 si ritrovano come unità singole, mentre le restanti sono associate con altri tratti di DNA in elementi compositi chiamati *BIME* (Bacterial Interspersed Mosaic Elements, Bachellier et al., 1997). Le sequenze *BIME* sono state identificate in diversi enterobatteri non necessariamente agli stessi siti cromosomali (Gilson et al., 1990).

Inoltre *BIME* localizzate nella stessa regione intergenica possono avere una struttura differente (Bachellier et al., 1997).

Sette differenti brevi motivi di DNA possono essere interspersi con gli elementi *PU* nelle sequenze *BIME* (Figura 1B). I segmenti *S* e *L* si trovano in corrispondenza della testa di elementi *PU* adiacenti (HIS, head internal sequences), i segmenti *l, s*, ed *r* a livello della coda (TIS, tail internal sequences). I due motivi esterni, *A* e *B* fiancheggiano la coda dell'ultima sequenza *PU* di un mosaico *BIME* (TES, tail external sequences) (Figura 1B).

In base al tipo di elemento *PU* presente e alla loro capacità di interagire con la proteina IHF (Integration Host Factor), è possibile dividere le sequenze *BIME* in due famiglie. Gli elementi *BIME-1* sono costituiti dai moduli *Z1* e *Y*, separati dal motivo *L* che contiene un sito di legame per IHF, e sono spesso fiancheggiati dai due motivi esterni *A* e *B*. Gli elementi *BIME-2*

sono composti da un numero variabile (da 2 a 12) di moduli *Z2* e *Y*, con il motivo *S* tra le loro teste e le sequenze *l* (oppure *s* o *r*) tra le loro code. In *E. coli* sono stati individuati 61 elementi *BIME-1* e 72 elementi *BIME-2* (Figura 1B). Le *BIME-1* tendono ad essere localizzate al 3' delle unità trascrizionali, mentre le *BIME-2* si trovano frequentemente tra geni di uno stesso operone.

Il ruolo delle *BIME* non é stato ancora completamente chiarito ed è stato suggerito che esse siano implicate in differenti processi, quali la stabilizzazione degli mRNA, la terminazione trascrizionale, il controllo della traduzione, il riarrangiamento genomico e l'organizzazione funzionale del cromosoma batterico.

È stato dimostrato che diverse proteine interagiscono con gli elementi *BIME*.

Le sequenze *PU* presenti in elementi *BIME-2* interagiscono *in vitro* con la DNA polimerasi I (Gilson et al., 1990). Il significato funzionale di questa interazione è ancora argomento di speculazione. Le *BIME* potrebbero servire come sito preferenziale di entrata per la DNA polimerasi I, fornendo un sito specifico di pausa per la reazione di polimerizzazione oppure giocando un ruolo nella fedeltà della replicazione.

La DNA girasi si lega alle sequenze *PU* presenti nelle *BIME-2* sia *in vitro* (Yang e Ames, 1988) che *in vivo* (Espéli e Boccard, 1997). L'interazione della DNA girasi con le *BIME-2* suggerisce un ruolo per questi elementi nel rilascio della tensione torsionale generata durante la replicazione e/o la trascrizione del DNA (Espéli e Boccard, 1997).

Circa 800 copie di una sequenza di 35 bp con una struttura di una palindroma imperfetta simile alle *REPs* è stata trovata nel genoma di *Pseudomonas putida*. La maggior parte di queste sequenze *REP*-simili sono localizzate tra geni trascritti convergentemente, e si ipotizza che il loro legame alla DNA girasi possa ridurre il grado di superavvolgimento positivo del DNA generato dalla trascrizione dei geni limitrofi (Aranda-Olmedo et al., 2002).

E' stato individuato un sito di legame per l'IHF (Integration Host Factor) nel motivo *L* che costituisce la porzione centrale delle *BIME-1*. Queste sequenze sono anche chiamate RIP (Repetitive IHF-binding Palindromic sequence) oppure RIB (Reiterative IHF-*BIME*).

L'IHF è una proteina simile agli istoni che è fondamentale per l'assemblaggio di strutture nucleoproteiche specializzate

coinvolte nella ricombinazione sito-specifica, nella replicazione, nella trasposizione o nella trascrizione (Nash, 1996).

È stato postulato che l'interazione della proteina IHF con gli elementi *BIME*-1 può facilitare o stabilizzare strutture nucleoproteiche che coinvolgono altre proteine che si legano agli elementi *PU*. In seguito al legame, la proteina IHF induce una forte curvatura di circa 140° degli elementi RIB; ciò potrebbe favorire il legame della DNA girasi e permettere la rimozione dei superavvolgimenti positivi generati dalla RNA polimerasi durante la trascrizione. In tal modo l'interazione della proteina IHF con gli elementi *BIME*-1 potrebbe influenzare la trascrizione degli operoni contenenti gli elementi *BIME* (Boccard e Prentki, 1993).

Le sequenze *BIME* possono agire non solo a livello di DNA, come componenti strutturali del cromosoma, ma anche a livello di RNA, come elementi di controllo dell'espressione di sequenze codificanti. Strutture a forcina, che si possono formare per appaiamento di segmenti complementari presenti nei trascritti, possono interferire con la terminazione, con la traduzione o con il processamento di sequenze codificanti co-trascritte (Gilson et al., 1986). Analizzando elementi *BIME* inseriti all'interno di operoni, è stato possibile dimostrare che le regioni codificanti localizzate

a monte hanno spesso livelli di espressione più elevati delle regioni codificanti localizzate a valle (Gilson et al., 1984; Newbury et al., 1987; Stern et al., 1988). Si pensa che tale effetto sia dovuto alla formazione nel trascritto di ripiegamenti in strutture stem-loop delle *PU* che proteggerebbero i segmenti di RNA a monte dall'attività degradativa delle esonucleasi batteriche con processività 3'-5' (Newbury et al., 1987). Questo effetto dipende dal numero di elementi *PU* presenti in una *BIME*, poiché un solo elemento *PU* non stabilizza un mRNA (Newbury et al., 1987), e dalla struttura del trascritto dove l'elemento *BIME* è inserito, poiché lo stesso elemento *BIME* ha un differente effetto sulla stabilizzazione degli mRNA a seconda dell'operone in cui è inserito (Ziemke and McCarthy, 1992). Sebbene l'aumento nell'espressione del gene a monte sia modesto, esso può essere biologicamente importante in alcuni operoni.

Le sequenze *BIME* possono influenzare la traduzione: la struttura secondaria alternativa adottata dalla regione intergenica potrebbe modificare l'accessibilità del sito di legame dei ribosomi inibendo la traduzione del gene a valle (Stern et al., 1988).

Recentemente, è stato suggerito per le sequenze *BIME* di *E. coli* un ruolo nell'attenuazione trascrizionale. Il fattore di terminazione Rho è richiesto in questo meccanismo di

attenuazione in associazione con le *BIME*. Queste ultime causerebbero una pausa nella progressione della RNA polimerasi e il fattore Rho potrebbe agire per terminare la trascrizione (Espéli O. et al., 2001).

Gli elementi *BOX* sono sequenze ripetute intergeniche presenti nel genoma di *Streptococcus pneumoniae*. Tali elementi sono tratti di DNA altamente conservato, con struttura modulare, composti da tre subunità (*boxA, boxB e boxC*), lunghe rispettivamente 59, 45 e 50 bp. La presenza di sequenze *BOX* a monte di geni implicati nei processi di virulenza suggerisce che queste sequenze possano essere regolatori dell'espressione genica in *S. pneumoniae* (Martin et al., 1992).

SCOPO DELLA TESI

Lo scopo della tesi è stato quello di definire la organizzazione strutturale e funzionale di due distinti tipi di sequenze ripetute presenti in genomi procariotici: gli elementi *nemis* e gli elementi *eric*.

Gli elementi *nemis* e *eric* rappresentano un componente cospicuo del genoma di *Neisseria. meningitidis* e *Yersinia enterocolitica*, rispettivamente.

Entrambi gli elementi sono caratterizzati dalla presenza di lunghe sequenze terminali ripetute invertite (TIRs) di 26-27 coppie di basi (bp), sono spesso localizzati vicino a geni cellulari, sono co-trascritti con le regioni codificanti vicine in mRNA e possono formare strutture a forcina molto stabili.

Abbiamo dimostrato che gli elementi *nemis* e *eric* possono entrambi influenzare il grado di espressione dei geni adiacenti a agendo a livello post-trascrizionale. I meccanismi attraverso cui questo si verifica sono alquanto distinti per i due tipi di sequenze intergeniche studiate.

Gli RNA messaggeri che contengono sequenze *nemis* (mRNA *nemis*+) sono processati a livello di strutture a forcina formate dall'appaiamento dei TIRs di *nemis* dall'endoribonucleasi III (RNasi III). Il processamento mediato dalla RNasi III può sia ritardare che facilitare il turnover dei trascritti contenenti gli

elementi *nemis*, e i livelli di espressione di trascritti omologhi *nemis*+ e *nemis*- in differenti ceppi di *N. meningitidis* sono difatto alquanto differenti.

Al contrario, le strutture a forcina formate da RNA *eric*, a seconda dell'orientamento e della posizione relativa dell'elemento nella molecola di mRNA, possono rallentare o aumentare la degradazione di trascritti *eric*+ da parte dell' apparato degradativo degli RNA batterici noto come degradosoma.

MATERIALI E METODI

Analisi bioinformatiche

Le sequenze *nemis* individuate nel genoma di *N. meningitidis* A Z2491 sono disponibili nel database GENEbank e sono state usate come *queries* in analisi di BLAST (Altschul et al. 1990) per trovare le sequenze omologhe presenti nel genoma di *N. meningitidis* MC58 e *N. gonorrheae* F1090.

Le sequenze *eric* di *E. coli*, disponibili al sito www.pasteur.fr/recherche/unites/pmtg/repet/intro.IRU.html, sono state usate come *queries* in analisi di BLAST per monitorare la distribuzione degli elementi *eric* in tutti i genomi procariotici sequenziati finora e in particolare per trovare le sequenze omologhe presenti nel genoma dei ceppi di *Y. pestis* CO92 e KIM e nel genoma del ceppo di *Y. enterocolitica* 8081.

L'identificazione degli elementi *eric* di *Y. enterocolitica* è stata possibile combinando le analisi BLAST con CLUSTALW (Thompson et al. 1994), un programma che consente di effettuare allineamenti multipli di sequenze disponibile al sito http://bioinfo.biogem.it/dbpcmuser/bioinfo/tools/winterf/clustalw-simple.php.

Le strutture secondarie degli elementi *nemis* ed *eric* sono state definite usando il programma Mulfold versione 2.0, che predice le strutture secondarie di RNA calcolandone il valore di energia

libera (Zuker, 1990) e descritte usando il programma loopDloop disponibile on-line al sito http://iubio.bio.indiana.edu.

Crescite batteriche

I ceppi di *N. meningitidis* e di *N. lactamica* (vedi Tabella 1) sono stati cresciuti nel brodo di coltura GC (per 1 litro: 7.5 g di DIFCO proteose, 0.5 g DIFCO solubile starci, 2 g K_2HPO_4, 0.5 g KH_2 PO_4, 2.5 g NaCl), addizionato con Polyvitox 1%, a 37°C in 5% CO_2.

Il ceppo Ye-161 (sierogruppo O8) di *Yersinia enterocolitica* isolato dall'Istituto di Igiene dell'Università di Amburgo, è stato fornito dalla Dr. Ida Luzzi dell'Istituto Superiore di Sanità di Roma. I ceppi Ye-24 (sierogruppo O8) e Ye-25 (sierogruppo O9) di *Yersinia enterocolitica* e il ceppo SS47 di *Yersinia kristensii* è stato fornito dalla Dr. Francesca Berlutti dell'Istituto di Igiene dell'Università La Sapienza di Roma.

I ceppi batterici di Yersinia sono stati cresciuti in LB arricchito (per 1 litro: 20 g di bactotriptone, 10 g di estratto di lievito, 10 g di NaCl) a 28°C in fase di crescita logaritmica.

Estrazione di DNA genomico

I batteri sono stati centrifugati a 3500 rpm per 15' a 4°C, e risospesi in 800 microlitri della soluzione 1 (per 1 litro: 9.9 g di 50 mM glucosio, 12.5 ml Tris 2M pH 7.5, 20 ml EDTA 0.5M pH 8). Per digerire la parete cellulare il campione è trattato prima con 1mg/ml di lisozima e, dopo 15' a temperatura ambiente, con una soluzione contenente 0.025M EDTA, 1% SDS e proteinasi K alla concentrazione finale di 0.2mg/ml. Il campione è stato incubato a 55°C per 2h. Si procede all'estrazione del DNA genomico prima con un volume di fenolo, poi con un volume di fenolo/cloroformio ed infine con un volume di cloroformio, centrifugando ad ogni passaggio a 15000 rpm per 15'. Il DNA estratto è stato precipitato in presenza di 0.3M sodio acetato pH 6.6 con 2.5 volumi di etanolo al 100%. Il DNA cromosomico ad alto peso molecolare è stato recuperato manualmente con una pipetta pasteur sterile, lavato in etanolo al 70% e trasferito in nuovi tubi. Dopo essiccazione, il DNA cromosomico recuperato è stato risospeso in TE (10mM TrisHCl pH 8, 1 mM EDTA pH 8).

Amplificazione di regioni genomiche mediante PCR

La reazione a catena della polimerasi (PCR) è stata usata per amplificare *in vitro* segmenti specifici del cromosoma, utilizzando come *primers* coppie di oligonucleotidi composti ciascuno da 50 residui. Gli oligonucleotidi sono stati opportunamente scelti in base ai dati di sequenza nucleotidica per amplificare dal genoma regioni di 450-500 bp contenenti sequenze ripetute analizzate. In alcuni casi i due *primers* utilizzati sono stati progettati in modo tale da contenere ciascuno un tratto di 25 bp omologo a segmenti di DNA delle regioni codificanti adiacenti agli elementi ripetuti. Le restanti 25 bp di ogni oligonucleotide contengono la sequenza del promotore riconosciuto dalla RNA polimerasi codificata dal batteriofago T7 o quella del promotore riconosciuto dalla RNA polimerasi codificata dal batteriofago Sp6.

I prodotti di amplificazione di DNA genomico (amplimeri) sono stati isolati mediante elettroforesi su gel di agarosio 1%. Dopo colorazione con etidio bromuro le bande di interesse sono state visualizzate mediante luce ultravioletta, rimosse con bisturi ed inserite in un tubo da dialisi. Gli amplimeri sono stati recuperati mediante elettroforesi. Il bromuro di etidio complessato con il

DNA è stato rimosso con due estrazioni fenoliche ed i campioni, dopo precipitazione con etanolo, sono stati risospesi in TE.

Estrazione di RNA totale da ceppi batterici

Le cellule batteriche sono state prelevate in fase di crescita logaritmica ad una densità ottica di 0,350 a 600 nm. Dopo centrifugazione a 3500 rpm per 15' a 4°C, i pellets batterici sono stati risospesi in 200 microlitri di TE. Per rompere la parete batterica è stato aggiunto lisozima alla concentrazione finale di 1mg/ml. I lisati batterici sono stati omogeneizzati in 500 microlitri di un buffer contenente guanidina isotiocianato per inattivare le ribonucleasi e consentire l'isolamento di mRNAs integri. Dopo aver aggiunto 350 microlitri di etanolo 100%, i campioni sono stati caricati su una colonnina di gel di silice "RNeasy" (Qiagen). Dopo lavaggi opportuni, RNA totale batterico è stato eluito con 80 microlitri di acqua sterile. Un decimo del campione è stato utilizzato per determinarne spettrofotometricamente la concentrazione (1 O.D. a 260nm = 40 microgrammi di RNA).

Analisi di RNA cellulari specifici

a) *Primer extension*

Per i saggi di primer extension sono stati utilizzati 20-30 μg di RNA per reazione. L'RNA è ibridato con 0,12 picomoli di oligonucleotide complementare marcato con ^{32}P in 30 μl di una soluzione contenente 500 mM Tris-HCl, pH 8,3, 80 mM $MgCl_2$, 300 mM KCl, 3 mM DTT. Dopo denaturazione per 5' a 95°C, i campioni sono incubati 10' a 68°C, 3 ore a 55°C e poi portati lentamente a temperatura ambiente. Ai campioni sono aggiunti 30 μl di una soluzione 500 mM Tris-HCl, pH 8,3, 80 mM $MgCl_2$, 300 mM KCl, 3 mM DTT contenente 0,8 mM desossinucleotidi trifosfati e 5 unità di RNasina. Dopo aggiunta di 100 unità di trascrittasi inversa SuperScript GIBCO, ogni campione è incubato 90' a 42°C. L'RNA è successivamente degradato mediante aggiunta di 10 μl NaOH 1N e incubando i campioni 15' a 55°C. dopo aggiunta di 10 μl HCl 1N ed estrazione con fenolo, i prodotti di reazione sono stati precipitati con etanolo, essiccati e risospesi in 5 μl di formamide contenente bromofenolo blu e xilene cianolo. Dopo denaturazione (5' 95°C), i campioni sono analizzati mediante elettroforesi.

b) "*RNase protection*"

i) Preparazione di sonde di RNA radioattive

Sono state allestite reazioni di trascrizione *in vitro* con RNA polimerasi del batteriofago Sp6, utilizzando come stampo gli amplimeri ottenuti per PCR (circa 50 nanogrammi), per creare sonde di RNA radioattive. Al termine della reazione, i campioni sono stati incubati per 1 ora a 37°C con 1 unità di DNasi RQ1 "ribonuclease-free" allo scopo di rimuovere completamente il DNA stampo per evitare eventuali interferenze nella successiva reazione di ibridazione. Dopo estrazione con un volume di miscela fenolo/cloroformio e successiva precipitazione in etanolo 100%, le sonde ottenute (riboprobes) sono state risospese in 20 microlitri di acqua sterile.

ii) Formazione di ibridi RNA/RNA resistenti alla degradazione da ribonucleasi

Per le reazioni di ibridazione sono stati incubati 5 microgrammi di RNA batterico totale e ~$4x10^4$ c.p.m. (conte per minuto) di riboprobe antisenso in 30 microlitri di 75% formammide, 25% HB 4X (per 1 ml: 80mM Tris pH 7.5, 4 mM EDTA, 1.6 M NaCl, 0.4% SDS) per 16-20 ore a 45°C. Dopo aggiunta di *Rnase digestion buffer* (per 1 ml: 10mM Tris pH 7.5, 5 mM EDTA, 300 mM NaCl, RNasi T1 alla concentrazione di 2 microgrammi/ml), i

campioni sono stati incubati per 60' a 33°C. La digestione con RNasi T1 è stata bloccata aggiungendo al campione proteinasi K (50 microgrammi/microlitro) ed incubandolo per 15' a 37°C. Dopo estrazione con una miscela di fenolo/cloroformio (rapporto 1:2), i campioni sono stati precipitati in etanolo, risospesi in 5 microlitri di buffer di caricamento (80% formammide, 0.025% blu di bromofenolo, 0.025% bromocianolo, 1X TBE), denaturati a 90°C per 5' e caricati su gel denaturante di poliacrilammide 6% - 8M urea.

c) RT-PCR

1 μg di RNA cellulare isolato da cellule batteriche arrestate in fase logaritmica è stato convertito in cDNA (DNA complementare) in 19 μl di una soluzione contenente 0,25 mM desossinucleosidi trifosfati, 3,2 unità/ml random primers, 8 mM DTT. Dopo denaturazione per 5' a 95°C, i campioni sono incubati 5' a temperatura ambiente; successivamente sono stati aggiunti 5 μl di reverse transcriptase buffer e 100 unità di trascrittasi inversa SuperScript GIBCO e i campioni sono stati incubati per 90' a 37°C. Il cDNA è stato successivamente co-amplificato utilizzando 0,7 picomoli di oligonucleotidi marcati con ^{32}P all'estremità 5'. I prodotti di PCR sono stati separati su

gel di poliacrilammide 6% urea 8M e l'intensità delle bande è stata quantizzata mediante PhosphorImager.

Preparazione di estratti cellulari batterici

Le cellule batteriche sono state cresciute fino a fase logaritmica. La crescita è stata arrestata a ~ 400 O.D. ed i batteri precipitati a 3500 rpm per 15' a 4°C. I "pellets" sono stati risospesi in ghiaccio in una soluzione contenente 10 mM Tris acetato pH 8, 14 mM magnesio acetato, 60 mM potassio acetato ed 1 mM DTT. Dopo rottura dei batteri alla French press, i campioni sono stati centrifugati a 15000 rpm per 30' a 4°C per eliminare membrane e pareti. La concentrazione proteica degli estratti è stata determinata con metodi colorimetrici (saggio BIORAD).

Saggi di degradazione *in vitro* di RNA

Le molecole di RNA radioattive sono state ottenute per trascrizione degli amplimeri con la RNA polimerasi del fago T7, in presenza di ATP, CTP e GTP (500 μM), UTP (250 μM) e ^{32}P UTP (25 μCi). I prodotti della trascrizione *in vitro* (circa $5x10^4$ c.p.m.) sono stati incubati con 3 μg di estratto cellulare per 20-40' a 37°C in un buffer contenente 25 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 60 mM KCl, 100 mM NH_4Cl, 0.1 mM DTT, 5%

glicerolo e 25 microgrammi/ml di RNA di lievito. Dopo la reazione di degradazione *in vitro* i campioni sono stati estratti con un volume di una miscela fenolo/cloroformio e quindi precipitati in 0.3 M sodio acetato pH 5.6 aggiungendo 20 microgrammi di tRNA e 2.5 volumi di etanolo 100%. Dopo lavaggio con etanolo al 70%, i *pellets* sono stati risospesi in 5 microlitri di buffer di caricamento (80% formammide, 0.025% blu di bromofenolo, 0.025% xilene cianolo, 1X TBE), denaturati a 90°C per 5' e caricati su gel denaturante di poliacrilammide 6% 8M urea. Mediante co-elettroforesi di un marker di peso molecolare noto è stato possibile determinare le dimensioni dei prodotti di degradazione relativamente alle dimensioni delle bande radioattive del marker.

Espressione della proteina ricombinante Rnasi III

Per clonare il gene *rnc* di *N. meningitidis* codificante per la RNasi III, due oligomeri di 31 bp (RNasi up: 5'-GAAAGTTGCTGCAGACGATGTTTTGAAACAG-3'; RNasi down: 5'-CGAATCAAGCTTGCCGCCTCATTTCTTTTTC-3' ; i residui sottolineati sono dedotti dalla sequenza di DNA del ceppo MC58 di *N. meningitidis*) sono stati utilizzati come primers per amplificare la regione codificante *rnc* dal DNA

genomico del ceppo B1940 di *N. meningitidis*. Il prodotto della amplificazione è stato digerito con gli enzimi di restrizione PstI e HindIII e clonati a valle del promotore del fago T7 tra i siti di restrizione PstI e HindIII del plasmide pRSETB (Invitrogen), che dirige la sintesi di trascritti, i quali sono tradotti da estratti cellulari di reticolociti (T&T Kit, Promega).

Saggi di band shift

RNA marcati con ^{32}P purificati da gel di poliacrilammide 6% 8M urea (circa 2000 c.p.m.) sono stati incubati con estratti cellulari batterici o con la proteina ricombinante RNasi III di *N. meningitidis*. I campioni sono stati incubati in 15 µl di 40mM Hepes (pH7,2), 20mM $(NH_4)_2SO_4$, 15 mM potassio acetato, 10% glicerolo, 50 mg/ml tRNA di lievito e 0,5 mg/ml BSA per 30' a 4°C e poi sono stati caricati su gel di poliacrilammide 5%. La corsa elettroforetica è stata condotta a 4°C per 2 ore a 23 mA in 0,5X Tris/borato buffer.

RISULTATI

Gli elementi *nemis*: un'abbondante famiglia di sequenze ripetute presenti nel genoma di *N. meningitidis*

Le Neisseriae sono batteri Gram negativi che includono diverse specie commensali e due specie patogene per l'uomo: *N. meningitidis* (o meningococco) e *N. gonorrhoeae* (o gonococco). Sebbene strettamente relazionate, le due specie patogene colonizzano diversi epiteli e causano malattie molto differenti.

N. gonorrhoeae è responsabile di infiammazioni del tratto urogenitale, *N. meningitidis* colonizza il nasofaringe ma può invadere l'epitelio, e raggiunto il flusso sanguigno causare setticemia e meningite attraversando la barriera ematoencefalica.

La determinazione della sequenza genomica completa dei ceppi di *N. meningitidis* di sierogruppo A Z2491 e di sierogruppo B MC58 ha fornito importanti informazioni per la caratterizzazione molecolare dei meningococchi.

Una caratteristica inattesa del genoma del meningococco è l'abbondanza e la varietà di sequenze di DNA ripetitivo non codogeniche, che rappresentano circa il 10% del genoma. Una quota cospicua di sequenze ripetitive non codogeniche (circa il 2% del genoma) è rappresentata dagli elementi *Correia* (Correia et al., 1986) o *nemis* (Neisseria Miniature Insertion Sequences, Mazzone et al., 2001). Esistono sia elementi *nemis* completi di

160 coppie di basi che elementi riarrangiati di 110 coppie di basi che differiscono dagli elementi completi per l'assenza di uno specifico segmento di 50 coppie di basi (Figura 2A). Gli elementi *nemis* terminano ad entrambe le estremità con il dinucleotide TA. La comparazione, in ceppi differenti, di siti cromosomali omologhi pieni e vuoti rivela che gli elementi *nemis* inducono la duplicazione del dinucleotide TA all'atto di integrazione genomica. Tutti gli elementi *nemis* sono caratterizzati da lunghe sequenze terminali ripetute invertite o TIRs di 26-27 coppie di basi, che hanno una struttura composita. Ogni TIR è costituita da un modulo esterno di 13 coppie di basi e uno interno di 10 coppie di basi, che può essere di tipo L o R, connessi da un segmento di DNA di 3 o 4 coppie di basi (TTG o CAAA) (Figura 2B). Il modulo di connessione determina sia la dimensione che il grado di complementarità delle TIRs di ciascun elemento. Inoltre poiché sono possibili due tipi di moduli interni e di due possibili moduli di connessione si possono avere quattro tipi di TIRs: 26L, 26R, 27L e 27R. Poiché ciascun *nemis* contiene due TIRs, sono possibili diverse combinazioni. Analisi bioinformatiche hanno permesso di identificare quattro principali sottofamiglie *nemis*: 26L/26R, 27L/27R, 27L/26R e 26L/27R. La sottofamiglia più ampia è rappresentata dagli elementi 26L/26R.

Nei due sierogruppi di *N. meningitidis* analizzati il numero di elementi *nemis* è simile (270 per il ceppo Z2491 e 250 per il ceppo MC58). In *N. gonorrhoeae*, invece, il numero di elementi *nemis* è notevolmente inferiore (110 per il ceppo F1090) sebbene le dimensioni genomiche di *N. meningitidis* e *N. gonorrhoeae* siano simili. Conseguentemente risulta che molti geni conservati nel meningococco e nel gonococco sono fiancheggiati da *nemis* soltanto in *N. meningitidis* ma non in *N. gonorrhoeae*.

In *N. meningitidis* il 70% dei *nemis* è localizzato vicino a geni cellulari. Oltre 150 elementi sono localizzati sia a monte che a valle delle ORFs ad una distanza inferiore a 100 bp (Figura 3). Tale distribuzione suggerisce che i *nemis* potrebbero avere funzioni regolatorie. Saggi di *primer extension* condotti su sette geni di *N. meningitidis* fiancheggiati all'estremità 5' da elementi *nemis* dimostrano che l'estremità 5' dei trascritti è localizzata all'interno dei TIRs di *nemis* (Mazzone et al 2001). Il dato è in perfetto accordo con la ipotesi che esista un processamento post-trascrizionale degli RNA contenenti *nemis*. Gli elementi *nemis* hanno la capacità di formare strutture stem-loop molto stabili con un ΔG a 37°C che varia per gli elementi completi e deleti da -52 a -39 kcal/mol, rispettivamente (Figura 2C). Strutture di questo tipo potrebbero essere il bersaglio di specifiche endoribonucleasi.

Trascritti *nemis*+ sono processati dai lisati cellulari di *E. coli* e di *N. lactamica* a livello di siti specifici

Per confermare questa ipotesi, abbiamo analizzato due regioni genomiche di *N. meningitidis* contenenti elementi *nemis*. Le regioni sono state amplificate mediante PCR (polymerase chain reaction) dal genoma del ceppo di *N. meningitidis* BL2 e inserite nel sito SmaI del vettore pGEM4Z tra i promotori fagici Sp6 e T7. Abbiamo così ottenuto i plasmidi pGEM-378, che contiene la parte iniziale della ORF NM1970 e un elemento *nemis* completo di 128 bp con TIRs lunghe 26 bp, e pGEM-417, che contiene la parte iniziale della ORF NM2107 e un elemento *nemis* riarrangiato di 81 bp con TIRs lunghe 27 bp.

I due plasmidi sono stati utilizzati come stampo per la trascrizione *in vitro* di RNA sintetizzati sia dalla RNA polimerasi del batteriofago Sp6 che dalla RNA polimerasi del batteriofago T7 in presenza di α-^{32}P UTP. I trascritti radioattivi ottenuti sono stati purificati da gel di poliacrilamide e incubati per 20' a 37°C con differenti quantità di estratti cellulari totali di *E. coli*. I prodotti di reazione sono stati separati mediante elettroforesi ad alta risoluzione e visualizzati nediante autoradiografia.

La degradazione dei substrati pGEM-378 T7 e pGEM-378 Sp6 (Figura 4) in prodotti di peso molecolare più basso indica che attività enzimatiche presenti negli estratti di *E. coli* sono in grado di mediare il processamento dei trascritti *nemis*+. Le bande a-f nell'autoradiogramma mostrato in figura 4 hanno la grandezza attesa in seguito al processamento dei trascritti in corrispondenza delle TIRs di *nemis*.

Il processamento dell'RNA potrebbe essere influenzato sia dalla sequenza dell'elemento *nemis* che dalla distanza tra le TIRs. Per verificare tale punto, i trascritti contenenti l'elemento *nemis* riarrangiato presente nel plasmide pGEM-417 (Figura 5) sono stati incubati con estratti cellulari di *E. coli*. Anche in questo caso, i principali prodotti di reazione (bande g-n nell'autoradiogramma in Figura 4) corrispondono a molecole di RNA generate dal taglio di trascritti *nemis*+ in posizioni specifiche nelle estremità TIRs.

Tali dati indicano che i trascritti contenenti *nemis*, che differiscono sia in grandezza che in sequenza, sono processati, almeno *in vitro*, con la stessa efficienza.

Incubando il trascritto pGEM-378 T7 con 0,5 μg (Figura 6, canale 4) e 1 μg (canale 5) di estratti cellulari del ceppo

apatogeno *N. lactamica* si osserva un profilo di degradazione analogo a quello ottenuto con gli estratti di *E. coli*.

Per definire con precisione i siti di taglio dei trascritti *nemis*+ da parte di attività enzimatiche presenti negli estratti di *E. coli* e *N. lactamica*, abbiamo condotto esperimenti di *primer extension* (Figura 7). Trascritti *in vitro* non radioattivi pGEM-378 T7 e pGEM-378 Sp6 (Figura 7A) sono stati dapprima incubati con 0,5 μg di estratti di *E. coli* (canale 1) e di *N. lactamica* (canale 2). I prodotti del processamento, b e f (vedi figura 4), così ottenuti sono successivamente ibridati con specifici oligomeri marcati con ^{32}P all'estremità 5' e complementari a regioni comprese tra i due TIRs (*bo* e *fo* rispettivamente). Questi sono stati poi elongati in presenza di desossiribonucleosidi trifosfati dalla trascrittasi inversa SuperScript II e separati su gel di poliacrilammide 8% urea 8M. La determinazione dei siti di taglio è stata resa possibile dalla co-elettroforesi di reazioni di sequenza dei plasmidi pGEM-378 e pGEM-417 ottenute utilizzando come molecola di innesco gli stessi oligonucleotidi usati come primer per l'analisi dei trascritti.

Lo stesso esperimento è stato fatto utilizzando trascritti non radioattivi pGEM-417 T7 e l'oligomero *lo* (Figura 7B), che è complementare a una regione a valle dell'elemento *nemis*.

L'oligonucleotide *lo* ibridizza sia con i prodotti di processamento l che con i prodotti di processamento l+n, permettendo in tal modo di identificare contemporaneamente entrambi i siti di taglio.

Come si osserva dalla Figura 7, il taglio degli elementi *nemis* avviene a livello di residui nucleotidici che si trovano all'interno dei TIRs (bande I-IV), anche se la posizione del taglio è diversa in dipendenza della grandezza e della sequenza dell'elemento *nemis* analizzato.

La endoribonucleasi III è responsabile del processamento degli elementi *nemis*

Per identificare le ribonucleasi coinvolte nel processamento di trascritti *nemis*+, sono stati effettuati saggi di degradazione *in vitro* con estratti cellulari ottenuti da ceppi di *E. coli* con mutazioni in geni che codificano specifiche ribonucleasi.

Il ceppo SK5003 ha un allele inattivo del gene *pnp*, che codifica la polinucleotide fosforilasi (PNP), un enzima con attività esonucleasica 3'-5' primariamente responsabile della degradazione di mRNA. Il ceppo SK5695, invece, possiede una mutazione temperatura-sensibile nel gene *rne1*, che codifica la RNasi E, che è l'unica endoribonucleasi fisiologicamente

coinvolta nel *turnover* generale degli mRNA batterici. Infine, il ceppo HT115 contiene una mutazione nel gene *rnc* e quindi produce una RNasi III inattiva. La figura 8 mostra l'assenza di bande di accumulo solo in corrispondenza del canale in cui il trascritto è incubato con estratti cellulari del ceppo HT115 (canale 5). Tali dati indicano che la RNasi III è l'enzima che riconosce specificamente la struttura a forcina formata dall'appaiamento delle sequenze TIRs ed è responsabile del suo processamento.

La RNasi III ricombinante di *N. meningitidis* è sufficiente per il processamento dei trascritti *nemis*+

Per avere una prova diretta del coinvolgimento della RNasi III nel processamento degli elementi *nemis*, abbiamo sintetizzato *in vitro* la RNasi III di *N. meningitidis*. A tale scopo abbiamo amplificato mediante PCR il gene *rnc* codificante la RNasi III di *N. meningitidis*. Il prodotto di PCR è stato inserito a valle del promotore A10 del fago T7 nel vettore di espressione pRSETB. Il plasmide ricombinante è stato usato come templato per dirigere l'espressione della proteina RNasi III in un sistema *in vitro* accoppiato di trascrizione-traduzione.

Il profilo di degradazione dei trascritti *nemis*+ ottenuti con la proteina ricombinante e con l'estratto cellulare di *N. lactamica* sono indistinguibili (Figura 9), indicando che la RNasi III è sufficiente sia per il riconoscimento che per il taglio dei trascritti *nemis*+.

Distribuzione degli elementi *nemis* nel genoma delle Neisseriae

L'ipotesi che le sequenze *nemis* hanno un ruolo nel controllo post-trascrizionale dell'espressione genica è validata dalla conservazione di questi elementi in differenti ceppi di *N. meningitidis*. Abbiamo monitorato la conservazione di una serie rappresentativa di membri della famiglia *nemis* in quindici ceppi di meningococchi e in tre ceppi apatogeni di *N. lactamica* (tabella 1). I 57 elementi *nemis* scelti e le rispettive ORFs che li fiancheggiano sono elencati in tabella 2.

Dieci nanogrammi di DNA genomico di ciascun ceppo sono stati amplificati mediante PCR, utilizzando cento nanogrammi di oligonucleotidi complementari a sequenze di DNA fiancheggianti ciascun elemento localizzate a 300-700 bp di distanza. Gli oligonucleotidi sono stati disegnati in base alla

conservazione di sequenza dei ceppi di *N. meningitidis* MC58 e Z2491.

I risultati di tale analisi sono illustrati in Figura 10. Le dimensioni dei prodotti di PCR ottenuti permettono di valutare se sequenze *nemis* sono presenti nelle regioni genomiche amplificate (siti pieni) o assenti (siti vuoti), e se in ceppi differenti si alternano allo stesso locus elementi completi ed elementi riarrangiati. I rettangoli grigio chiaro e grigio scuro in Figura 10 rappresentano gli elementi *nemis* normali e riarrangiati, rispettivamente. Guardando la distribuzione dei grigi si nota subito che, ad esempio, ai siti corrispondenti agli elementi 5, 7, 28, 48, 49, 50 e 55, elementi *nemis* completi e riarrangiati si alternano nei diversi ceppi di *N. meningitidis*. Un'ulteriore causa di variabilità è rappresentata dall'esistenza di regioni genomiche anomale (indicate in Figura 10 dal simbolo #), che non corrispondono in grandezza né a siti pieni né a siti vuoti.

Trentuno dei cinquantasette elementi *nemis* sono conservati in tutti i ceppi di meningococchi analizzati. La localizzazione della maggior parte degli elementi *nemis* analizzati è abbastanza conservata in ceppi di *N. meningitidis* che appartengono a differenti sierogruppi. Gli elementi sono maggiormente conservati in ceppi appartenenti allo stesso "lineage". Ad

esempio *nemis* 19, 20 e 42 non sono presenti nei ceppi appartenenti al complesso ET-5; *nemis* 19, inoltre, è assente anche nei ceppi appartenenti al cluster L1; *nemis* 55 è assente nei ceppi appartenenti al lineage 4; *nemis* 51 è assente nei ceppi del lineage 4 e nei ceppi del cluster L1 (Figura 10).

Il numero dei siti *nemis*+ presenti in *N. lactamica* é sorprendentemente basso. Solo tre elementi sono conservati in tutti i ceppi di *N. lactamica*; venti elementi sono conservati in uno o due ceppi ma ben trentaquattro sono assenti in tutti i ceppi. Questi dati suggeriscono che gli elementi *nemis* sono circa tre volte meno abbondanti in *N. lactamica* rispetto a *N. meningitidis* (vedi istogramma in Figura 10). In accordo alle analisi *in silico*, gli elementi *nemis* sono sottorappresentati anche nel ceppo F1090 di *N. gonorrhoeae*. Questo potrebbe suggerire che gli elementi *nemis* sono comparsi in cellule ancestrali prima della separazione delle Neisseriae in specie patogene e non patogene e poi si sono diffusi in maniera selettiva solo nei meningococchi.

Trascritti mini-*nemis* contenenti TIRs di differente tipo non sono processati dalla RNasi III

I livelli di espressione della maggior parte dei geni elencati in Tabella 2 potrebbero essere influenzati dal processamento degli

elementi *nemis*, data la breve distanza che li separa dagli elementi. Per determinare se tutte le varianti strutturali degli elementi *nemis* (26/26, 27/27 e 26/27) sono processate con la stessa efficienza nello stesso contesto di RNA, abbiamo utilizzato i costrutti artificiali di DNA mini-*nemis* (Figura 11). I corrispondenti trascritti, che hanno lo stesso contesto di RNA ma hanno TIRs dello stesso tipo o di tipo differente separati da 10 nucleotidi (nt), sono stati utilizzati come substrati in esperimenti di degradazione *in vitro*. Strutture di RNA a forcina formate da TIRs lunghi 26 nt (mini-26/26a) o 27 nt (mini-27/27) sono efficientemente processate (canali 1 e 5 rispettivamente). Al contrario trascritti contenenti TIRs di 26 e 27 nt (mini-26/27) sono completamente resistenti al processamento da parte della RNasi III (canali 2). Nei trascritti mini-26/27 la struttura di RNA a doppia elica formata dai TIRs è in parte distrutta da un loop interno asimmetrico, che potrebbe inibire il taglio da parte della RNasi III. Questo dato è in disaccordo con i dati presenti in letteratura, in quanto substrati che presentano loops interni sono generalmente tagliati dalla RNasi III di *E. coli* a un lato o a entrambi i lati della struttura a doppia elica (Li et al 1993, Nicholson, 1999). Inoltre stems di RNA di 11 bp sono efficientemente processati dalla RNasi III di *E. coli* e i trascritti

mini-26/27 possono formare uno stem di RNA lungo 13 bp (vedi Figura 11). Abbiamo costruito anche tre varianti del costrutto mini-26/26a. I cambi nucleotidici introdotti nei costrutti mini-26/26b e mini-26/26c alterano la composizione della sequenza dei TIRs, ma non l'energia della struttura di RNA a doppia elica formata dai TIRs. Tali cambi non hanno effetto sul processamento ad opera della RNasi III (canali 3 e 4 rispettivamente).

Questi dati indicano che il grado di complementarietà dei TIRs è il maggiore determinante per i trascritti *nemis*+ e che la struttura a forcina, piuttosto che la sequenza, è cruciale per il processamento da parte della RNasi III.

Interazione tra la RNasi III e i trscritti *nemis*+

Abbiamo utilizzato i trascritti mini-26/26a e mini-26/27 anche come sonde in esperimenti di *band shift* (Figura 12). Entrambe le classi di trascritti formano complessi RNA-proteina con la stessa mobilità, dopo incubazione dei trascritti *nemis*+ radioattivi con la proteina RNasi III ricombinante (canale 2 e 7) o con estratti cellulari di *N. lactamica* (canale 3 e 8). Da ciò si può dedurre che il legame della RNasi III con gli RNA *nemis*+ avviene anche se non si ha il processamento degli RNA *nemis*+. Inoltre si

osservano differenze nella grandezza dei complessi RNA-proteina quando i trascritti *nemis*+ sono incubati con la proteina ricombinante (canale 2 e 3) o con gli estratti cellulari (canali 7 e 8). Questo dato è in accordo con la nozione che l'attività della RNasi III negli estratti cellulari è predominantemente associata a complessi ad alto peso molecolare.

Quando le sonde di RNA sono incubate con estratti cellulari di *E. coli*, complessi RNA-proteina si formano solo con gli estratti cellulari del ceppo rcn+ (canale 4 e 9) ma non con gli estratti cellulari del ceppo rcn- (canale 5 e 10), che è deficiente della RNasi III a causa dell'inserzione del trasposone Tn10 nella regione codificante del gene della RNasi III.

I dati suggeriscono che i complessi RNA-proteina sono RNasi III-dipendenti e si formano solo in presenza di una RNasi III funzionale; inoltre altri fattori cellulari fanno parte di questi complessi ma non interagiscono direttamente con gli RNA *nemis*+.

Analisi dei trascritti *nemis*+ in vivo

Per verificare i risultati ottenuti con i trascritti artificiali mini-*nemis*, abbiamo amplificato diverse regioni del cromosoma di *N. meningitidis* contenenti elementi *nemis* strutturalmente differenti,

utilizzando coppie di oligonucleotidi che includono i promotori fagici T7 e Sp6. I prodotti di PCR così ottenuti hanno permesso di sintetizzare *in vitro* trascritti successivamente utilizzati sia come substrati per saggi di degradazione sia come sonde antisenso in esperimenti di protezione degli ibridi RNA/RNA dall'attacco di ribonucleasi (saggio di *RNase protection*). In particolare abbiamo analizzato due regioni cromosomali che contengono due elementi *nemis* 26/27 (*nemis* 23 e *nemis* 28). In entrambi i casi non si osserva degradazione dei trascritti senso *nemis*+ quando questi sono incubati con estratti cellulari di *N. lactamica* (Figura 13A).

Abbiamo inoltre condotto saggi di *RNase protection* su RNA cellulari isolati da cellule di *N. meningitidis* BL859 arrestate in fase logaritmica precoce. Le sonde antisenso radioattive formano ibridi RNA/RNA con i corrispettivi mRNA cellulari che risultano protetti dalla azione degradativa dell'RNasi T1, una endoribonucleasi che degrada selettivamente gli RNA a singolo filamento. Mediante esperimenti, abbiamo osservato l'accumulo di bande discrete che hanno una grandezza corrispondente a specie di RNA non processate (Figura 13B).

Questi dati confermano che gli elementi *nemis* 26/27 non sono processati dalla RNasi III anche *in vivo*.

Alcuni elementi *nemis* 26/26 sono resistenti al taglio operato dalla RNasi III. Analisi di *RNase protection* mostrano che i trascritti *nemis*+ contenenti *nemis* 54, un elemento 26/26, non sono processati *in vivo* (Figura 13C, pannello a sinistra). Al contrario i trascritti *nemis*+ contenenti *nemis* 47, un altro elemento 26/26, sono processati *in vivo* (Figura 13C, pannello a destra). Comunque non si può escludere che i trascritti *nemis*+ contenenti *nemis* 54 sono processati dalla RNasi III ma non riusciamo a vedere i prodotti del loro processamento a causa della loro rapida degradazione.

Questi dati suggeriscono che lo specifico contesto di RNA nel quale i trascritti *nemis*+ si trovano potrebbe influenzare il taglio da parte della RNasi III; ciò comporta che il numero di trascritti *nemis*+ non processati non può essere stabilito a priori sulla base di informazioni bioinformatiche.

Avendo identificato ceppi di *N. meningitidis* nei quali sequenze *nemis* sono assenti in specifiche regioni genomiche (vedi Figura 10), abbiamo misurato i livelli di espressione di alcuni trascritti *nemis*+ e *nemis*- mediante saggi di *primer extension* (Figura 14). RNA totale isolato da ceppi di *N. meningitidis* che differiscono per la presenza o per l'assenza allo stesso locus di un elemento *nemis* é stato ibridizzato con specifici oligomeri marcati con ^{32}P

all'estremità 5' complementari alle regioni codificanti. Come si vede in Figura 14A i livelli di espressione dei trascritti *nemis*+ codificanti la ORF1585 sono comparabili in ceppi *nemis*+ e in ceppi *nemis*-. Al contrario, RNA processati a livello delle sequenze *nemis* codificanti la ORF1244 sono 5 volte meno abbondanti dei trascritti non processati presenti nel ceppo *nemis*- (Figura 14B). Invece i livelli di espressione dei trascritti corrispondenti alla ORF811 sono più elevati in ceppi *nemis*+ e non sono misurabili in ceppi *nemis*- (Figura 14C). Le differenze quantitative che emergono da questi esperimenti sono in accordo con i dati presenti in letteratura: infatti, è noto che il taglio ad opera della RNasi III può aumentare la stabilità o accelerare la degradazione di specifici trascritti (Nicholson, 1999; Coburn, 1999).

Gli elementi *eric*: una famiglia di sequenze ripetute presenti nel genoma di *Y. enterocolitica*

Durante il periodo di dottorato mi sono interessata anche a un'altra famiglia di sequenze ripetute, gli elementi *eric*, presenti nel genoma delle Yersiniae.

Il genere *Yersinia* comprende 10 specie, di cui *Y. pestis* e *Y. enterocolitica* sono quelle più conosciute come patogene per l'uomo. La *Y. pestis* causa due forme di peste, la peste bubbonica (più comune) e la peste polmonare, entrambe caratterizzate da un alto tasso di mortalità. Le infezioni causate da *Y. pestis* sono zoonosi e l'uomo è solo un ospite accidentale. La malattia si diffonde attraverso il morso di pulci, per contatto diretto con tessuti infetti, o da persona a persona per inalazione di aerosol infettanti (peste polmonare). *Y. enterocolitica* infetta l'uomo per ingestione di alimenti o acqua contaminata e causa principalmente gastroenteriti. La virulenza della *Y. enterocolitica* è associata a specifici sierotipi. I sierotipi più frequentemente isolati in Europa, Africa, Giappone e Canada sono O3 ed O9. *Y. enterocolitica* di sierogruppo O8 è stata identificata come agente primario di infezione negli Stati Uniti (Bottone, 1997).

Recentemente il genoma del ceppo 8081 di sierogruppo O8 di *Y. enterocolitica* è stato completamente sequenziato al "Sanger Centre". Gli elementi *eric* sono membri di una famiglia di DNA ripetuto presenti in basso numero di copie in *Escherichia coli* (Sharples e Lloyd, 1990; Hulton et al., 1991). Mediante un approccio bioinformatico, nel nostro laboratorio è stata condotta una ricerca mirata all'identificazione di sequenze *eric* in altri genomi procariotici. Il risultato di questa analisi conferma che un basso numero di copie (15-25 elementi) di sequenze *eric* è presente in differenti enterobatteri, quali *Shigella flexneri*, *Klebsiella pneumoniae*, *Salmonella typhi* e *Salmonella typhimurium*. Sorprendentemente, gli elementi *eric* sono particolarmente abbondanti in differenti specie di *Yersinia*, nelle quali essi costituiscono dallo 0.45% (*Y. pestis*) all'0,7% (*Y. enterocolitica*) dell'intero genoma.

Utilizzando come sonde sequenze *eric*, sono state condotte analisi di BLAST sul genoma di *Y. enterocolitica* e di *Y. pestis*. I risultati ottenuti hanno portato all'identificazione di 247 e di 167 elementi *eric* completi nel genoma di *Y. enterocolitica* di *Y. pestis*, rispettivamente. Da ora in avanti mi riferirò selettivamente agli elementi di *Y. enterocolitica*.

In base alle dimensioni è possibile suddividere gli elementi in tre sottofamiglie (Figura 15): elementi normali (N, 126-127 bp), elementi lunghi (L, 192-217 bp) ed elementi corti (S, 68-80 bp). Gli elementi corti differiscono dagli elementi normali per la assenza di uno stesso tratto di DNA centrale. Gli *eric* lunghi sono elementi normali contenenti inserzioni localizzate in siti specifici. In base al motivo inserito è possibile distinguere due tipi di elementi lunghi (tipo I e tipo II). Dei 247 elementi presenti nel genoma di *Yersinia enterocolitica* 20 sono di tipo *short*, 39 di tipo *long* e 188 hanno la lunghezza degli elementi base.

Gli elementi *eric* posseggono TIRs di 26-27 bp, relativamente conservati. La comparazione di sequenze omologhe in *Y. pestis* e *Y. enterocolitica* ha permesso di identificare siti pieni (*eric*+) e siti vuoti (*eric*-). In questi ultimi, al posto di *eric* è presente il dinucleotide TA, una sequenza che è ripetuta alle estremità della sequenza consensus degli elementi *eric* (Figura 16). Ciò indica che gli *eric* sono (o sono stati) sequenze di DNA mobile la cui inserzione è associata alla specifica duplicazione del dinucleotide TA.

Abbiamo definito la posizione relativa di tutti gli elementi *eric* nel genoma di *Y. enterocolitica*. Nella maggior parte dei casi, gli elementi *eric* si ritrovano in singola copia a breve distanza da

sequenze codificanti (0-100 bp dalle triplette di inizio e/o fine; Figura 17).

Ciò fa ipotizzare che gli elementi *eric*, come già osservato per gli elementi *nemis*, possano regolare, a livello post-trascrizionale, l'espressione di geni adiacenti. L'ipotesi è suffragata dalla osservazione che RNA eric possono similmente ripiegarsi in strutture di tipo stem-loop (vedi Fig. 24).

Per verificare se e come le strutture secondarie di RNA formate da elementi *eric* sono bersaglio di eventi specifici di processamento, sono stati condotti saggi di degradazione *in vitro* in cui trascritti *eric*+ sono stati incubati con estratti cellulari di *Y. enterocolitica*. La degradazione del substrato in prodotti di peso molecolare più basso indica chiaramente che attività enzimatiche presenti negli estratti di *Y. enterocolitica* mediano il processamento dei trascritti *eric*+. Ciò nonostante, il numero e le dimensioni delle bande non permettono di ricostruire facilmente un profilo degradativo. Benché i trascritti *eric*+ possano formare strutture secondarie molto simili a quelle formate da trascritti *nemis*+, essi non costituiscono un bersaglio preferenziale per l'attività endonucleolitica della RNasi III.

Analisi dei trascritti *eric*+ in *Y. enterocolitica*

Data la peculiare distribuzione genomica (vedi figura 17), molti elementi *eric* sono presumibilmente co-trascritti con le sequenze codificanti vicine. Per verificare questa ipotesi, abbiamo monitorato mediante *primer extension* l'espressione di quattro geni, *lpdA*, *uncE*, *cheW* e *trpB*, fiancheggiati all'estremità 5' da un elemento *eric* nel ceppo sequenziato di *Y. enterocolitica* 8081. Mediante analisi di PCR, abbiamo verificato che nel ceppo batterico di *Y. enterocolitica* Ye-161 le regioni analizzate hanno la stessa configurazione presente nel ceppo 8081. RNA totale estratto da Ye-161 é stato ibridato con specifici oligomeri marcati con ^{32}P all'estremità 5' complementari alle regioni codificanti dei quattro geni. Nel caso dei geni *lpdA* e *uncE*, abbiamo osservato bande di estensione le cui grandezze corrispondono a trascritti che comprendono l'elemento *eric* (Figura 18, pannelli A e B). Almeno per *lpdA*, questo dato è in accordo con dati presenti in letteratura. Infatti in *E. coli* l'elemento *eric* nella regione genica *ace-lpdA* è conservato e trascritti *lpdA* che includono l'elemento *eric* sono stati identificati (Cunningham et al., 1998).

Al contrario, nel caso dei geni *cheW* e *trpB* si osservano bande di estensione multiple corrispondenti a classi di trascritti le cui

estremità 5' cadono all'interno degli elementi *eric* (Figura 18, pannelli C e D).

L'orientamento degli elementi *eric* rispetto ai codoni di termine delle ORFs adiacenti è asimmetrico e influenza la degradazione dei trascritti

Gli elementi *eric* possono essere diversamente inseriti (orientamento A o B) relativamente alla direzione della trascrizione dei geni adiacenti.

Analizzando sia l'orientamento in cui sono inseriti che la distanza relativa che separa gli elementi *eric* dai codoni di termine delle regioni codificanti limitrofe, abbiamo osservato che esiste una stretta correlazione tra l'orientamento degli elementi *eric* e la distanza dalle triplette di termine localizzate a monte (Figura 19). Elementi *eric* localizzati a distanza di 0-5 bp dalle ORFs a monte sono tutti inseriti nell'orientamento B. Quando la distanza fra gli elementi *eric* e il codone di stop è maggiore di 5 bp l'orientamento delle sequenze *eric* risulta essere prevalentemente quello A.

Questa singolare distribuzione è certamente non casuale ed è fortemente suggestiva di ruoli funzionali alternativi esercitati da elementi *eric.*

L'ipotesi più verosimile è che eventi di processamento mediati dall'elemento *eric* possano influenzare il grado di stabilità di trascritti omologhi ai geni localizzati a monte degli elementi *eric*, e che elementi *eric* inseriti in orientamento opposto possano differentemente concorrere a tale processo. Per verificare questa ipotesi abbiamo misurato i livelli di espressione di coppie di geni separati da elementi *eric* inseriti nell'orientamento A o B mediante saggi di RT-PCR (reverse transcriptase polymerase chain reaction, Figura 20). Tutti gli elementi *eric* analizzati sono localizzati tra ORFs trascritte nella stessa direzione (elementi DU).

L'RNA totale isolato da cellule di Ye-161 arrestate in fase logaritmica è stato convertito in cDNA (DNA complementare) utilizzando come primer corti oligomeri costituiti da sei nucleotidi di sequenza casuale. Il cDNA ottenuto è stato successivamente co-amplificato utilizzando oligonucleotidi gene-specifici marcati con ^{32}P all'estremità 5'. Per evitare l'effetto plateau, per ciascuna coppia di geni analizzati abbiamo determinato il numero di cicli in cui persiste una relazione lineare tra la quantità di RNA stampo e il prodotto finale di amplificazione. I prodotti di PCR sono stati separati su gel di

poliacrilammide 6% urea 8M, e l'intensità delle bande è stata quantizzata mediante PhosphorImager.

Come si evince dalla Figura 21, il rapporto trascritti geni a valle/ geni a monte cambia notevolmente in base all'orientamento dell'elemento *eric*. Eccetto per i trascritti *aceA* e *aceB*, i quali si accumulano allo stesso livello, quando tra i due geni è presente un elemento *eric* nell'orientamento B i trascritti del gene a valle sono in media circa quattro volte più abbondanti rispetto ai trascritti del gene a monte. Al contrario, quando è presente un elemento *eric* nell'orientamento A, il rapporto trascritti geni a valle/ geni a monte diminuisce in media al valore di 1.5 (Figura 21). I dati sono stati confermati utilizzando RNA estratti da ceppi di Yersiniae privi di sequenze *eric* in specifiche regioni genomiche. In Ye-161 i geni *cheA* e *cheW* sono separati da un elemento *eric* di 127 bp e i trascritti *cheW* sono circa quattro volte più abbondanti dei trascritti *cheA*. Utilizzando RNA estratti dalla specie apatogena *Y. kristensii* abbiamo verificato che i trascritti *cheA* e *cheW* si accumulano allo stesso livello (Figura 22). Analisi di sequenza hanno mostrato che nel ceppo di *Y. kristensii* utilizzato un elemento *eric* non si è mai inserito nella regione genica *cheA-cheW* ritrovandosi una copia intatta del sito di integrazione TA.

Nel ceppo Ye-161, un elemento *eric* di 202 bp si trova tra i geni *argB* e *argH* ed è inserito nell'orientamento B a 0 bp dal codone di termine del gene *argB*. In questo ceppo, i trascritti *argH* sono circa nove volte più abbondanti dei trascritti *argB*. Invece nel ceppo Ye-25, tra i geni *argB* e *argH* è presente un elemento *eric* di 127 bp che è inserito nell'orientamento A a 10 bp dal codone di termine del gene *argB*. In questo caso il rapporto dei trascritti *argH/argB* si riduce drasticamente, passando da nove a tre (Figura 23). Ciò conferma come la stabilità dei trascritti può essere influenzata sia dall'orientamento che dalla distanza dell'elemento *eric* dalle ORFs.

Il significato della distribuzione asimmetrica dei due orientamenti degli elementi *eric* non è immediato.

Gli RNA corrispondenti a elementi *eric* inseriti nell'orientamento A e nell'orientamento B si possono ripiegare in strutture secondarie relativamente simili, che hanno un'energia libera comparabile (Figura 24A).

Quando gli elementi *eric* sono lontani dai codoni di stop, indipendentemente dal loro orientamento, gli RNA corrispondenti possono liberamente ripiegarsi nelle strutture a forcina mostrate.

Invece RNA di elementi *eric* localizzati vicino ai codoni di stop (nella maggior parte dei casi sono elementi inseriti in orientamento B), non possono correttamente ripiegarsi. Questo è dovuto al fatto che un ribosoma in temporanea posizione di stallo alla tripletta di stop copre la sequenza TIR prossimale dell'elemento *eric,* dal momento che un ribosoma "copre" un tratto di circa 30 nucleotidi di un RNA messaggero (Steitz., 1969). Questo implica che gli RNA corrispondenti agli elementi *eric* inseriti vicino ai codoni di stop non possono formare la struttura a forcina mostrata in figura 24B. Questo determina lo smascheramento di una sequenza conforme al consenso RAUUW (R= A, G; W= A, U), che rappresenta il sito di taglio preferenziale per la RNasi E (Ehretsmann et al 1992), la endoribonucleasi responsabile della degradazione degli mRNA nei batteri (Babitzke e Kushner, 1991, Figura 24C)

Strutture a forcina formate dagli elementi *eric* potrebbero rallentare la progressione delle esonucleasi batteriche in direzione 3'-5', stabilizzando i trascritti a monte degli elementi. Quando però gli elementi *eric* sono localizzati vicini ai codoni di stop si potrebbe avere il taglio da parte della RNasi E, i trascritti *eric*+ diventano più accessibili all'attacco da parte delle esonucleasi 3'-5' e di conseguenza la degradazione dei trascritti a

monte dell'elemento *eric* risulta aumentata; questo modello spiega perfettamente i dati riportati in Figura 21.

Un supporto a tale modello è fornito da esperimenti in cui abbiamo disaccoppiato i processi di trascrizione e traduzione.

Le cellule del ceppo Ye-161 sono state cresciute a 28°C e in fase logaritmica è stato aggiunto il cloramfenicolo, un antibiotico che blocca la sintesi proteica. Le cellule sono state prelevate dopo 30 minuti dall'aggiunta dell'antibiotico e l'RNA cellulare è stato isolato. Abbiamo misurato, mediante saggi di RT-PCR, i livelli di espressione di due coppie di geni separati da elementi *eric* inseriti in entrambi gli orientamenti sia in cellule di controllo che in cellule alle quali è stato aggiunto l'antibiotico (Figura 24D).

Nel caso dei geni *cheA-cheW* separati da un elemento *eric* nell'orientamento B, il rapporto cheW/cheA si riduce di un fattore circa tre. Quando invece l'elemento *eric* è nell'orientamento A, come nel caso dei geni *panB-panC*, il rapporto trascritti geni a valle/ geni a monte non varia.

DISCUSSIONE

Oltre ai trasposoni e alle sequenze di inserzione, nei genomi procariotici sono presenti differenti classi di DNA ripetitivo la cui funzione è ignota o solo parzialmente conosciuta.

In questa tesi abbiamo descritto due famiglie di DNA ripetitivo batteriche, gli elementi *nemis* e gli elementi *eric*, che influenzano l'espressione dei geni adiacenti agendo entrambi come sequenze di controllo a livello di RNA.

La prima famiglia analizzata è rappresentata dalle sequenze *nemis*, un'abbondante classe di elementi presenti nel genoma delle Neisseriae. In base alla lunghezza, gli elementi *nemis* vengono distinti in elementi *long* (154-158 bp) ed elementi *short* (104-108 bp), privi di un segmento centrale di 50 bp. Entrambi i tipi sono caratterizzati da lunghe sequenze terminali invertite (TIRs) di 26-27 bp. Gli elementi *nemis* sono co-trascritti con i geni cellulari adiacenti in mRNA che sono processati a livello di strutture a forcina formate dall'appaiamento delle loro strutture terminali o TIRs (Figura 4-6). Saggi di degradazione *in vitro* suggeriscono che l'enzima coinvolto nel processamento dei trascritti *nemis*+ è la RNasi III (Figura 8 e 9). La RNasi III di *E. coli* è una endoribonucleasi specifica per l'RNA a doppio filamento che contiene due domini funzionali. La porzione C-terminale contiene un dominio di legame all'RNA a doppio

filamento (dominio dsRBD), la porzione N-terminale rappresenta il dominio catalitico o DSRM (Conrad et al., 2002). Entrambi i domini sono conservati nella RNasi III di *N. meningitidis*, che presenta una omologia del 49% e una similarità del 66.7% con la RNasi III di *E. coli*. La RNasi III clonata di *N. meningitidis* è sufficinete per riconoscere e processare correttamente i trascritti *nemis*+ (Fig. 9).

Nei batteri, la RNasi III regola principalmente il processamento dei trascritti precursori degli rRNA (Bram et al., 1980) e occasionalmente partecipa alla maturazione di specifiche classi di mRNA (Regnier et al., 1990).

L'abbondanza e la distribuzione degli elementi *nemis* suggerisce che la RNasi III funga da regolatore globale in *N. meningitidis*, modulando la maturazione di numerosi classi di mRNA.

Dall'analisi genomica condotta su ceppi di meningococchi di differenti sierotipo e origine geografica risulta che gli elementi *nemis* rappresentano un componente stabile del genoma dei meningococchi, mentre sono sottorappresentati nei ceppi della specie apatogena *N. lactamica*. La maggior parte delle regioni intergeniche analizzate risulta *nemis*+ in *N. meningitidis* ma non in *N. lactamica* (Figura 10). Probabilmente eventi di trasposizione hanno determinato l'espansione selettiva di questa

famiglia di elementi di DNA nel genoma dei meningococchi. Tenendo conto che scambi genetici avvengono con elevata frequenza tra *N. meningitidis* e *N. lactamica* che colonizzano la stessa nicchia ecologica (Linz et al., 2000), la marcata asimmetria nella distribuzione degli elementi *nemis* tra ceppi patogeni e apatogeni può fare ipotizzare che i meningococchi possono avere un selettivo vantaggio nel mantenere elementi *nemis* in specifici siti cromosomali. Geni fiancheggiati dagli elementi *nemis* (vedi Tabella 2) codificano per regolatori trascrizionali, proteine regolatorie, proteine coinvolte nella patogenesi. L'ipotesi che il *processing* dei trascritti *nemis*+ mediato dalla RNasi III possa essere importante per la vita del meningococco come microrganismo patogeno è corroborata dall'osservazione che la RNasi III è un gene essenziale per lo sviluppo della batteremia nei ratti (Sun et al., 2000).

Come illustrato in Figura 11, i trascritti che contengono *nemis* con TIRs dello stesso tipo (elementi 26/26 o elementi 27/27) sono efficientemente tagliati dalla RNasi III; al contrario, i trascritti che contengono *nemis* con TIRs differenti (elementi 26/27) formano delle strutture a forcina che sono riconosciute dalla RNasi III ma non sono processate dall'enzima. Cambi di basi che alterano la composizione della sequenza ma non il grado

di complementarietà dei TIRs, non hanno effetto sul processamento ad opera della RNasi III.

Il grado di complementarietà dei TIRs, piuttosto che la sequenza specifica, è cruciale per il *processing* dei trascritti *nemis*+ sia *in vitro* che *in vivo* (Figure 11 e 13).

I substrati dell'RNasi III sono, generalmente, strutture secondarie di RNA a doppio filamento di 15-20 bp. L'RNasi III taglia la struttura riconosciuta a livello dello *stem* con due incisioni asimmetriche, ciascuna su uno dei filamenti che formano lo *stem*, sfalsate fra loro di due nucleotidi (Zhang et al., 1997).

La lunghezza minima degli *stems* riconosciuti dalla RNasi III di *E. coli* é di 11 bp (Nicholson, 1999). Benché trascritti di elementi *nemis* 26/27 possono formare uno stem di RNA di 13 bp, essi non sono processati dalla RNasi III di *N. meningitidis*.

Gli elementi *nemis* 26/27 rappresentano solo un quarto della famiglia *nemis*. Ma il numero di trascritti *nemis*+ refrattari al taglio da parte della RNasi III può essere più ampio di quello previsto da analisi *in silico*, perché il contesto di RNA può influenzare il taglio delle sequenze *nemis* da parte della RNasi III *in vivo* (Figura 13C).

Sia le strutture a forcina tagliate che quelle non tagliate interagiscono con la RNasi III di *N. meningitidis* (Figura 12). Di

conseguenza il legame degli elementi *nemis* non processati potrebbe regolare finemente i livelli della RNasi III, mediante la titolazione dell'enzima.

Il processamento degli elementi *nemis* mediato dalla RNasi III può ritardare o facilitare il turnover dei trascritti contenenti gli elementi *nemis*. Difatti analisi di *primer extension* dimostrano che *in vivo* i livelli di espressione dei trascritti *nemis*+ e *nemis*- in ceppi di *N. meningitidis* sono differenti, indicando che (Figura 14).

Non tutte le sequenze di RNA ripiegabili in strutture a doppia elica sono riconosciute e processate dalla RNasi III. È questo il caso degli elementi *eric*, una famiglia di DNA ripetuto analizzata in dettaglio solo in *E. coli* e *S. typhimurium* (Sharples e Lloyd, 1990; Hulton et al., 1991). In questi genomi gli elementi *eric* sono presenti in basso numero di copie (20-30). Al contrario gli elementi *eric* rappresentano un componente cospicuo del genoma di *Y. enterocolitica* (circa 0,7%). e sono caratterizzati da TIRs di 26-27 coppie di basi, che possono formare strutture a forcina molto stabili.

Gli elementi *eric* e *nemis* sono simili per molti aspetti:

a) entrambe le famiglie includono membri lunghi e corti che differiscono per la presenza/assenza di un segmento centrale di circa 50 bp, ma che conservano sequenze terminali intatte

b) l'integrazione dei due tipi di elementi è associata alla duplicazione della stessa sequenza bersaglio (target) costituita dal dinucleotide TA

c) entrambi i tipi di elementi presentano sequenze terminali invertite (TIRs) lunghe 26-27 bp e parzialmente complementari che possono ripiegarsi, a livello di RNA, in strutture secondarie.

Un ulteriore punto in comune tra gli elementi *eric* e *nemis* è rappresentato dalla loro distribuzione genomica. Come gli elementi *nemis*, anche gli elementi *eric* presenti in *Y. enterocolitica* sono localizzati in regioni intergeniche e sono spesso inseriti a breve distanza (0-100 bp) dalle triplette di inizio e/o di fine di regioni codificanti.

La formazione di strutture a forcina dovute all'appaiamento dei TIRs degli elementi *eric* può interferire con i livelli di espressione dei geni co-trascritti con gli elementi *eric* con un meccanismo che può aumentare o diminuire la degradazione degli mRNA.

La RNasi E è il principale enzima coinvolto nella degradazione degli mRNA batterici e taglia gli RNA a singola elica in una regione ricca in AU. Nei batteri la degradazione di mRNA avviene ad opera del degradosoma (Carpousis et al., 1994), un complesso multienzimatico nel quale sono presenti l'endoribonucleasi E (RNasi E), la polinucleotide fosforilasi (PNPasi), la RNA elicasi DEAD-box (RhlB) e una enolasi (Carpousis, 2002). Le proteine DnaK, GroEL e la polifosfato chinasi (PPK), sono presenti nel degradosoma in quantità substechiometriche. La funzione di queste ultime e dell'enolasi rimane ancora da chiarire.

La RNasi E introduce un taglio endonucleotidico nell'mRNA rendendolo suscettibile all'azione degradativa delle esonucleasi 3'-5' (PNPasi e ribonucleasi II). L'attività di entrambi enzimi è impedita dalla formazione di strutture secondarie negli mRNA.

Le sequenze ripetute intergeniche REPs (Repetitive Extragenic Palindromis sequences) formano una significativa frazione del genoma di *Escherichia coli* e *Salmonella typhimurium.* Molte REPs sono co-trascritte con geni cellulari, e si ripiegano nella molecola di mRNA in strutture a forcina che bloccano la progressione delle esonucleasi, stabilizzando in tal modo le regioni trascritte localizzate a monte. Due enzimi, la RhlB e la

poly(A) polymerase (PAP), sono richiesti *in vivo* per facilitare la degradazione degli mRNA contenenti REPs (Khemici et al., 2004).

Anche gli elementi *eric* formano strutture a doppia elica abbastanza stabili. Inaspettatamente il 95% degli elementi *eric* localizzati vicino ai codoni di stop dei geni sono inseriti nell'orientamento B. L'orientamento e la relativa posizione delle sequenze *eric* sono critici nel regolare i livelli dei trascritti complementari ai geni a monte e a valle.

Nei procarioti la trascrizione e la traduzione sono fisicamente accoppiate. Conseguentemente, il *folding* degli elementi *eric* inseriti vicino ai codoni di stop è inibita poiché il TIR prossimale è occupato dai ribosomi. In tal modo un sito target per la RNasi E presente nel TIR distale degli elementi *eric* inseriti nell'orientamento B non è mascherato dalla formazione di strutture a doppia elica e i segmenti della regione trascritta localizzati a monte sono suscettibili all'azione degradativa delle esonucleasi 3'-5'. Al contrario, i trascritti che contengono gli elementi *eric* inseriti nell'orientamento A possono ripiegarsi liberamente in una struttura a forcina relativamente stabile (Figura 24B). Questi elementi presumibilmente funzionano come

barriere intergeniche, come le REPs di *E. coli*, impedendo la progressione in direzione 5'-3' del degradosoma.

Le sequenze *nemis* ed *eric* hanno un'origine comune. Entrambe le sequenze sono (o sono state) elementi di DNA mobile che specificamente inducono la duplicazione del dinucleotide TA in seguito all'integrazione nel genoma.

Gli elementi *nemis* ed *eric* ricordano per diversi aspetti sequenze di DNA mobile chiamate MITEs (Miniature Inverted-repeat Transposable Elements), identificate in un gran numero di genomi di invertebrati e vertebrati (Feschotte et al., 2002).

I MITEs sono un particolare tipo di elementi trasponibili non autonomi, cioè non codificanti trasposasi, che conservano tuttavia sequenze in *cis* necessarie per la trasposizione.

I MITEs sono sequenze di ridotte dimensioni (meno di 600 bp) caratterizzate dalla presenza di sequenze terminali invertite (TIRs). Sono comuni nei genomi degli eucarioti ma sono stati trovati anche in alcuni archeobatteri (Brugger et al., 2002). L'unica altra famiglia di elementi MITE eubatterici è rappresentata dagli elementi RUP di *S. pneumoniae*. Nel caso di RUP la mobilizzazione sembra essere mediata dalla trasposasi codificata dalla IS630-Spn1 (Oggioni et al., 1999).

Analisi *in silico* non permettono di identificare trasposasi capaci di mobilizzare gli elementi *nemis* ed *eric*.

È interessante notare come membri di queste due famiglie di MITEs, distribuite in batteri Gram negativi molto distanti nella scala evolutiva, si sono evoluti come substrati per le principali ribonucleasi endocellulari. Analisi future potrebbero portare alla scoperta di ulteriori MITEs batterici che possono essere coinvolti nella regolazione del metabolismo degli mRNA.

BIBLIOGRAFIA

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J., 1990.
Basic local alignment search tool.
J Mol Biol. 215(3), 403-410.

Aranda-Olmedo I., Tobes R., Manzanera M., Ramos J.L., Marques S.,2002.
Species-specific repetitive extragenic palindromic (REP) sequences in Pseudomonas putida.
Nucleic Acids Res. 30, 1826-33.

Babitzke P., Kushner S.R., 1991.
The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of Escherichia coli.
Proc Natl Acad Sci 88, 1-5.

Bachellier S., Clément J.M., Hofnung M., Gilson E., 1997.
Bacterial Interspersed Mosaic Elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence.
Genetics 145, 551-562.

Bachellier S., Clément J.M., Hofnung M., 1999.

Short palindromic repetitive DNA elements in enterobacteria: a survey.

Res. Microbiol. 150, 627-639.

Boccard F., Prentki P., 1993.

Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units.

EMBO J. 12, 5019-5027

Bottone E.J., 1997.

Yersinia enterocolitica: the charisma continues.

Clin. Microbiol. Rev. 10, 257-276

Bram RJ., Young RA., Steitz JA., 1980.

The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E.coli*.

Cell 19, 393-401.

Brugger K., Redder P., She Q., Confalonieri F., Zivanovic Y., Garrett R.A., 2002.
Mobile elements in archaeal genomes.
FEMS Microbiol Lett., 206, 131-141.

Carpousis A.J., Van Houwe G., Ehretsmann C., Krisch HM., 1994.
Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation.
Cell. 76, 889-900

Carpousis A.J., 2002.
The Escherichia coli RNA degradosome: structure, function and relationship to other ribonucleolytic multienzyme complexes.
Biochem. Soc. Trans. 30, 150-155.

Coburn G.A., Mackie G.A., 1999.
Degradation of mRNA in Escherichia coli: an old problem with some new twists.
Prog Nucleic Acid Res Mol Biol. 62, 55-108.

Conrad C., Rauhut R., 2002.

Ribonuclease III: new sense from nuisance.

The International Journal of Biochemistry & Cell Biology 34, 116-129.

Correia FF., Inouye S., Inouye M., 1986.

A 26-base-pair repetitive sequence specific for Neisseria gonorrhoeae and Neisseria meningitidis genomic DNA.

J Bacteriol. 167, 1009-1015.

Cunningham L., Georgellis D., Green J., Guest JR., 1998.

Co-regulation of lipoamide dehydrogenase and 2-oxoglutarate dehydrogenase synthesis in E. coli: characterisation of an ArcA binding site in lpdA promoter.

FEMS Microbiology Letters 169, 403-408.

Ehretsmann C.P., Carpousis A.J., Krisch H.M., 1992.

Specificity of Escherichia coli endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site.

Genes Dev. 6, 149-159.

Espéli O., Boccard F., 1997.

In vivo cleavage of *Escherichia coli* BIME-2 repeat by DNA gyrase: genetic characterization of the target and identification of the cut site.

Mol. Microbiol. 26, 767-777.

Espeli O., Moulin L., Boccard F.,2001.

Transcription attenuation associated with bacterial repetitive extragenic BIME elements.

J Mol Biol. 314,375-386.

Feschotte C., Jiang N., Wessler S.R., 2002.

Plant transposable elements: where genetics meets genomics.

Nat. Rev. Genet. 3, 329-341.

Gilson E., Clément JM., Brutlag D., Hofnung M., 1984.

A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*.

EMBO J., 1417-1421.

Gilson E., Rousset J.P., Clément J.M., Hofnung M., 1986.
A subfamily of *E. coli* Palindromic Units implicated in transcription termination?
Ann. Inst. Pasteur/ Microbiol. 137B, 259-270.

Gilson E., Perrin D., Hofnung M., 1990.
DNA polymerase I and a protein complex bind specifically to *E. coli* Palindromic Units highly repetitive DNA: implications for bacterial chromosome organization.
Nucleic Acids Res. 18, 3941-3952.

Gilson E., Bachellier S., Perrin S., Perrin D., Grimont PAD., Grimont F., Hofnung M., 1990.
Palindromic Units highly repetitive DNA sequences exhibit species specificity within enterobacteriaceae.
Res. Microbiol. 141, 1103-1116.

Higgins CF., Ferro-Luzzi Ames G., Barnes WM., Clément JM., Hofnung M., 1982.
A novel intercistronic regulatory element of prokaryotic operons.
Nature 298, 760-762.

Hulton C.S., Higgins C.F., Sharp P.M., 1991.
ERIC sequences: a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria.
Mol Microbiol. 5, 825-834.

Khemici V., Carpousis A.J., 2004.
The RNA degradosome and poly(A) polymerase of Escherichia coli are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers.
Mol. Microbiol. 51, 777-790.

Li H.L., Chelladurai B.S., Zhang K., Nicholson A.W., 1993.
Ribonuclease III cleavage of a bacteriophage T7 processing signal. Divalent cation specificity, and specific anion effects.
Nucleic Acids Res. 21, 1919-1925.

Linz B., Schenker M., Zhu P., Achtman M., 2000.
Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*.
Mol. Microbiol. 36, 1049-1058.

Martin B, Humbert O, Camara M, Guenzi E, Walker J, Mitchell T, Andrew P, Prudhomme M, Alloing G, Hakenbeck R, 1992.
A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae.
Nucleic Acids Res. 20(13), 3479-3483.

Mazzone M., De Gregorio E., Lavitola A., Pagliarulo C., Alifano P., Di Nocera P.P., 2001.
Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*.
Gene 278, 211-222.

Mahillon J., Leonard C., Chandler M., 1999.
IS elements as constituents of bacterial genomes.
Res Microbiol. 150, 675-687.

Nash H.A., 1996
The HU and IHF proteins: accessory factors for complex protein-DNA assemblies.
In Regulation of Gene Expression in *Escherichia coli* (Lin E.C.C. &Lynch A.S., eds), 149-179.

Newbury S.F., Smith N.H., Higgins C.F., 1987.
Differential mRNA stability controls relative gene expression within a polycistronic operon.
Cell 51, 1131-1143.

Nicholson A.W., 1999.
Function, mechanism and regulation of bacterial ribonucleases.
FEMS Microbiology Reviews 23, 371-390.

Oggioni M.R., Claverys J.P., 1999.
Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*.
Microbiology 145, 2647-2653.

Regnier P., Grunberg-Manago M., 1990.
RNase III cleavages in non-coding leaders of Escherichia coli transcripts control mRNA stability and genetic expression.
Biochimie. 72, 825-834.

Sharples G.J., Lloyd R.G., 1990.

A novel repeated DNA sequence located in the intergenic regions of bacterial chromosome.

Nucl. Acids Res. 18, 6503-6508.

Steitz J.A., 1969.

Nucleotide sequences of the ribosomal binding sites of bacteriophage R17 RNA.

Cold Spring Harb Symp Quant Biol. 34, 621-630.

Stern M.J., Ferro-Luzzi Ames G., Smith N.H., Robinson E.C., Higgins C.J., 1984.

Repetitive extragenic palindromic sequences: a major component of the bacterial genome.

Cell 37, 1015-1026.

Stern M.J., Prossnitz E., Ferro-Luzzi Ames G., 1988.

Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium:* involvement of REP sequences.

Mol. Microbiol. 2, 141-152.

Sun Y.H., Bakshi S., Chalmers R., Tang C.M., 2000.
Functional genomics of *Neisseria meningitides* pathogenesis.
Nat. Med. 6, 1269-1273.

Thompson J.D., Higgins D.G., Gibson T.J., 1994.
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.
Nucleic Acids Res. 22, 4673-4680.

Yang Y., Ames G.F.L., 1988.
DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences.
Proc. Natl. Acad. Sci. *USA*, 85, 8850-8854.

Zhang K, Nicholson A.W., 1997.
Regulation of ribonuclease III processing by double-helical sequence antideterminants.
Proc Natl Acad Sci U S A. 94, 13437-13441.

Ziemke M., McCarthy J.E.G, 1992.

The control of mRNA stability in *Escherichia coli*: manipulation of the degradation pathway of the polycistronic *atp* mRNA.

Biochim. Biophys. Acta. 1130, 297-306

Zuker M., 1989.

On finding all suboptimal foldings of an RNA molecule.

Science 244, 48-52.

TABELLE

Ceppo	Sierogruppo	Gruppo epidemiologico	Origine	Fonte'
N. meningitidis				
BF2	B	ET-37 complex	Italia	a
93/4286	C	ET-37 complex	Norvegia	b
NGP165	B	ET-37 complex	Norvegia	b
FAM18	C	ET-37 complex	USA	web
BZ169	B	ET-5 complex	Olanda	b
H44/76	B	ET-5 complex	Norvegia	b
MC58	B	ET-5 complex	Scozia	web
205900	A	Subgroup IV-1	Italia	b
Z2491	A	Subgroup IV-1	Gambia	web
BL859	B	Lineage 3	Italia	c
BS845	B	Lineage 3	Italia	c
BL892	B	Lineage 3	Francia	d
BF9	B		Italia	a
B1940	B		Germania	e
BL947	B		Francia	d
NGF26	B		Norvegia	b
NGE31	B		Norvegia	b
NGH15	B		Norvegia	b
N. lactamica				
21			Francia	d
411			Francia	d
4627			Francia	d

Tabella 1: Ceppi di Neisseria utilizzati per analisi genomiche.

ORF	Funzione	*nemis*						Funzione	ORF
11	murA, UDP-N-acetylgl. carboxyvinyltransf.	26	u	**1**	d	310	transmembrane transport protein	**12**
51	pilU, twitching motility protein	100	d	**2**	u	30	integral membrane protein	**50**
89	pykA, pyruvate kinase	173	d	**3**	u	34	outer membrane protein	**88**
189	integral membrane protein	33	u	**4**	u	222	conserved hypothetical protein	**188**
194	alanine symporter protein	135	d	**5**	u	35	gidA, regulatory protein	**193**
295	ffh, signal recognition particle protein	62	d	**6**	d	114	dsbA, thiol:disulphide interchange protein	**294**
329	pilF, type IV assembly protein	46	u	**7**	d	289	conserved hypothetical protein	**328**
353	hypothetical protein	67	u	**8**	d	-4	hypothetical protein	**352**
380	anaerobic transcriptional regulator	35	u	**9**	u	122	hemN, coproporphyrinogen III oxidase	**379**
388	integral membrane transport protein	28	u	**10**	u	204	ABC transporter ATP-binding protein	**387**
390	mapA, maltose phosphorylase	45	u	**11**	d	35	galM, aldose 1-epimerase	**389**
534	membrane protein	45	u	**12**	u	390	transmembrane hexose transporter	**535**
614	putative oxidoreductase	13	d	**13**	d	76	amtB, probable ammonium transporter	**615**
619	conserved hypothetical protein	62	d	**14**	d	21	phosphoglycolate phosphatase	**620**
671	putative malate oxidoreductase	63	d	**15**	u	330	tetraacyldisaccharide 4'-kinase	**672**
698	unknown protein	40	d	**16**	u	40	tryptophan synthase, beta subunit	**699**
699	tryptophan synthase, beta subunit	14	d	**17**	d	59	IgA endopeptidase	**700**
723	rplT, 50S ribosomal protein L20	114	d	**18**	u	125	pheS, phenylalanyl-tRNA synthetase a chain	**724**
785	recB, exodeoxyribonuclease V beta chain	28	u	**19**	d	50	unknown protein	**786**
811	murB, UDP-N-acetylenolpyruvoylgl. reduct.	53	u	**20**	d	19	transmembrane efflux protein	**812**
823	adk, adenylate kinase	320	d	**21**	u	49	pyrF, orotidine 5'-phosphate decarboxylase	**824**
845	PhoH-related protein	39	u	**22**	d	-57	LPS biosynthesis related protein	**846**
885	dnaB, putative replicative DNA helicase	126	d	**23**	u	26	FimT, fimbrial protein	**886**
956	sucB, dihydrolipoamide succinyltransf. E2 comp.	49	d	**24**	u	209	lpdA, dihydrolipoamide dehydrogen. E3 comp	**957**
1241	cca, tRNA nucleotidyltransferase	21	d	**25**	u	96	hypothetical protein	**1242**
1244	rpe, ribulose phosphate epimerase	23	u	**26**	d	184	hypothetical protein	**1245**
1257	site-specific DNA methylase, pseudogene	18	u	**27**	d	123	unknown protein	**1258**
1287	iron sulphur binding protein	17	u	**28**	d	101	nrdB, ribonucleoside-diphosphate reductase	**1288**
1302	hip, integration factor beta subunit	112	d	**29**	d	84	putative transcriptional regulator	**1303**
1331	uvrB, excinuclease ABC subunit B	124	u	**30**	d	99	prc, carboxy-terminal processing protease	**1332**
1421	nifR3 protein	150	u	**31**	u	62	ATP-dependent RNA helicase	**1422**
1433	putative lipoprotein	50	u	**32**	d	3	phospholipase D family protein	**1434**
1438	hypothetical iron-sulphur protein	57	d	**33**	d	19	purE, phosphoribosylaminoimidazole carboxyl.	**1439**
1474	possible tautomerase	90	u	**34**	u	99	possible periplasmic protein	**1475**
1558	dgk, diacylglycerol kinase	25	u	**35**	d	132	gshB, glutathione synthetase	**1559**
1563	transcriptional regulator [GntR-family]	102	u	**36**	d	315	hypothetical protein	**1564**
1584	unknown protein	88	d	**37**	u	34	transcriptional regulator [MarR-family]	**1585**
1650	transcriptional regulator [AsnC-family]	43	u	**38**	u	174	alr, alanine racemase	**1651**
1669	hemO, haem utilisation protein	62	u	**39**	u	66	integral membrane protein	**1670**
1695	hypothetical protein	-89	d	**40**	d	139	hypothetical protein	**1699**
1706	unknown protein	-5	d	**41**	u	27	integral membrane ion transporter	**1707**
1710	gdhA, glutamate dehydrogenase	96	d	**42**	d	38	transcriptional regulator [GntR family]	**1711**
1711	lrp, transcriptional regulator [GntR family]	42	u	**43**	u	234	integral membrane protein	**1712**
1716	mtrC, membrane fusion protein	57	u	**44**	u	192	transcriptional regulator [mtrR family]	**1717**
1861	prmA, ribosomal protein L11	18	u	**45**	d	62	accC, acetyl-CoA carboxylase	**1862**
1877	prolyl endopeptidase	28	u	**46**	d	65	argA, acetylglutamate synthase	**1876**
1883	hypothetical protein	101	d	**47**	d	30	ferric siderophore receptor protein	**1882**
1897	leuS, leucyl-tRNA synthetase	159	d	**48**	u	23	drg, type II restriction endonuclease	**1896**
1918	fabD, malonyl CoA-acyl c. p. transacylase	74	u	**49**	d	-44	integral membrane protein	**1917**
1933	atpC, ATP synthase epsilon chain	146	d	**50**	u	103	glyQ, glycyl-tRNA synthetase alpha chain	**1932**
1953	sspA, stringent starvation protein A	42	u	**51**	d	86	hypothetical protein	**1954**
1956	rpmE, 50S ribosomal protein L31	81	d	**52**	u	66	cad, cadmium resistance protein	**1955**
1957	putative acetyltransferase	21	u	**53**	u	109	rpmE, 50S ribosomal protein L31	**1956**
2056	rpsI, 30S ribosomal protein S9	105	d	**54**	d	43	transcriptional regulator [metR family]	**2055**
2066	tldD, regulatory function	46	u	**55**	u	237	conserved hypothetical protein	**2067**
2071	thiG, thiamine biosynthesis protein	14	u	**56**	d	43	unknown protein	**2070**
2104	mafA, adhesin	45	u	**57**	d	61	pyrH, uridylate kinase	**2103**

Tabella 2: Elementi nemis mostrati in Figura 10

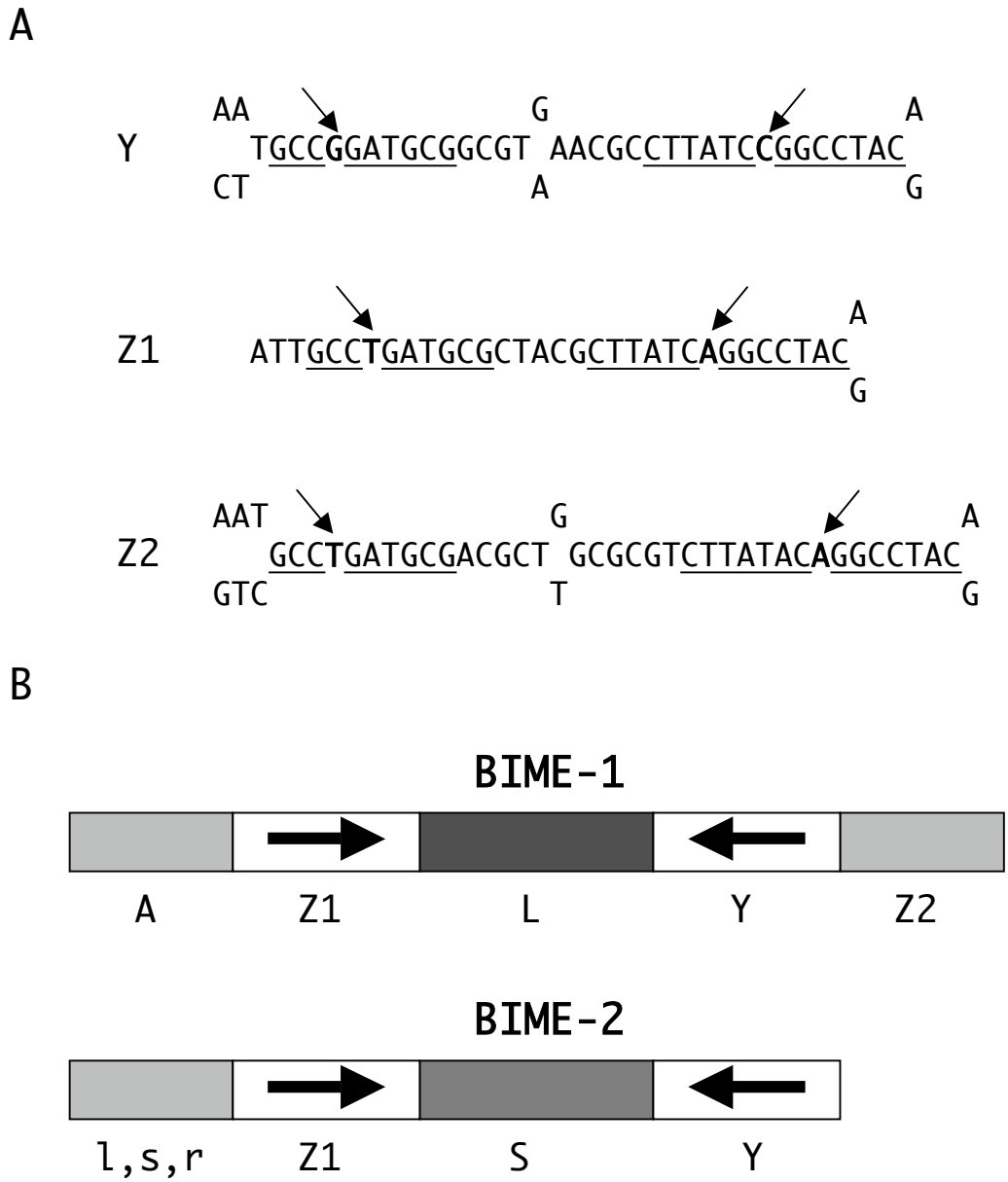


Figura 1. Organizzazione strutturale degli elementi *BIME* in *E. coli*. A) I nucleotidi mutati alla settima e trentaduesima posizione sono indicati dalle frecce. Le basi sottolineate indicano i nucleotidi conservati nei tre tipi di elementi PU. B) Nel mosaico *BIME* gli elementi *PU* hanno orientamento invertito.

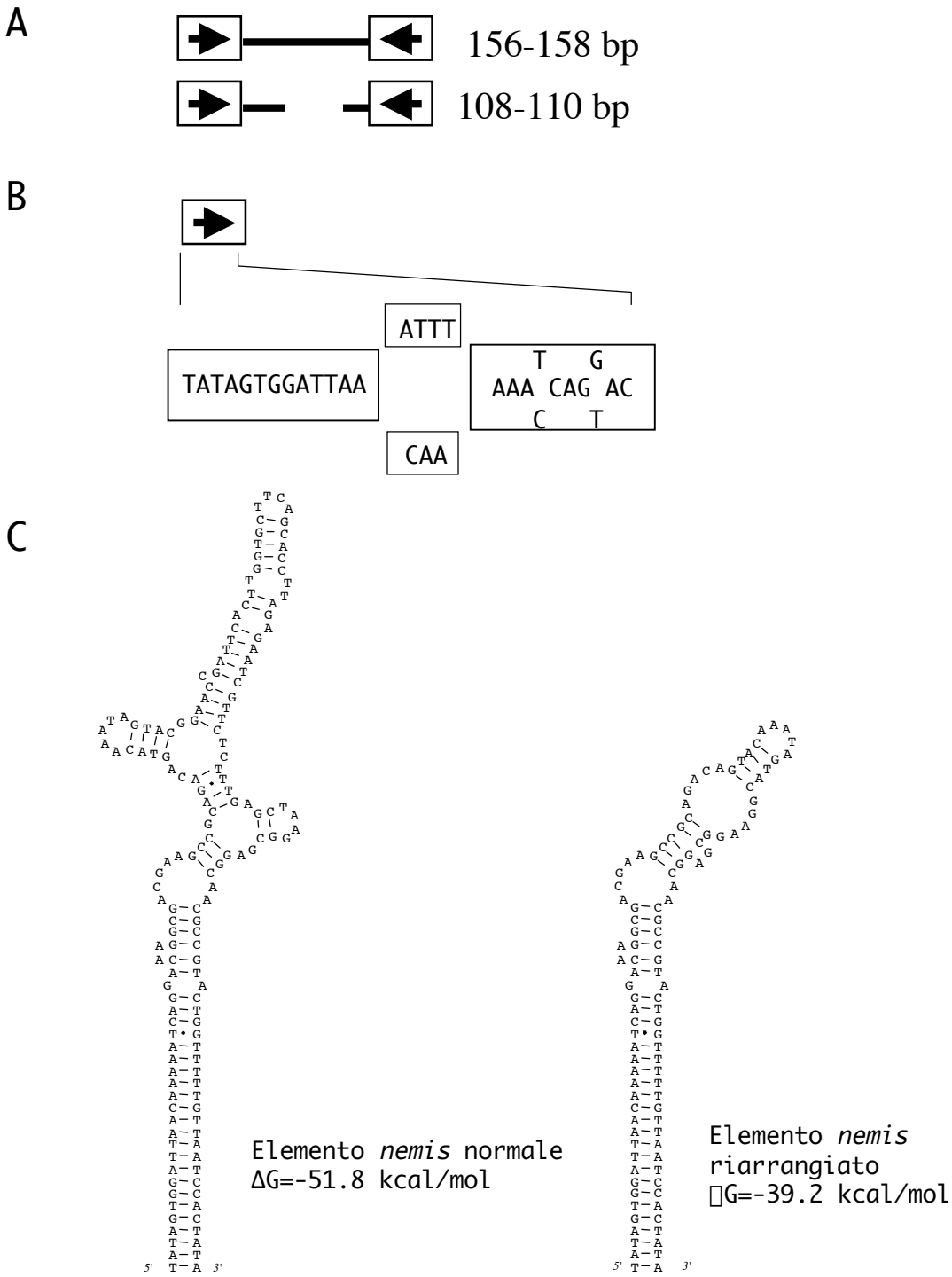


Figura 2. Organizzazione strutturale degli elementi *nemis*. A) Classi di elementi *nemis*. B) Struttura delle TIRs. C) Strutture secondarie a forcina di elementi *nemis*.

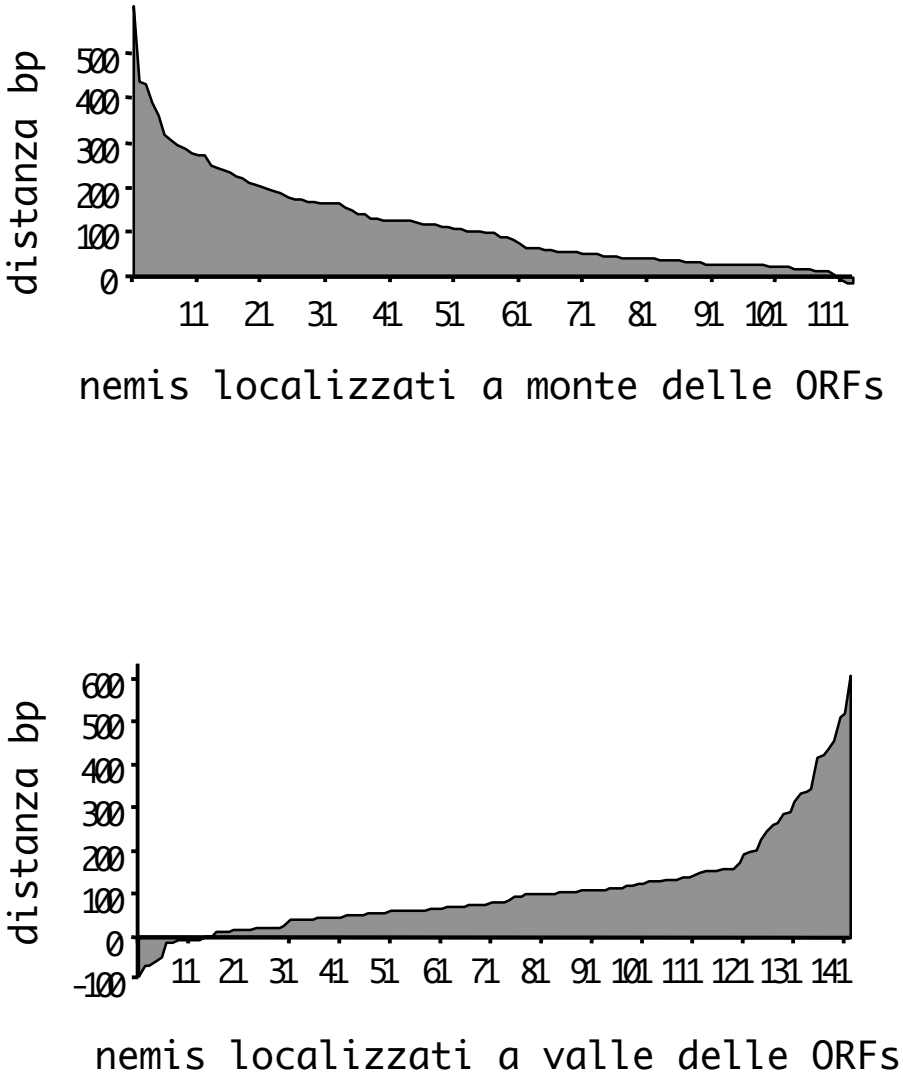


Figura 3. Distanza degli elementi *nemis* di *N. meningitidis* dalle regioni codificanti adiacenti.

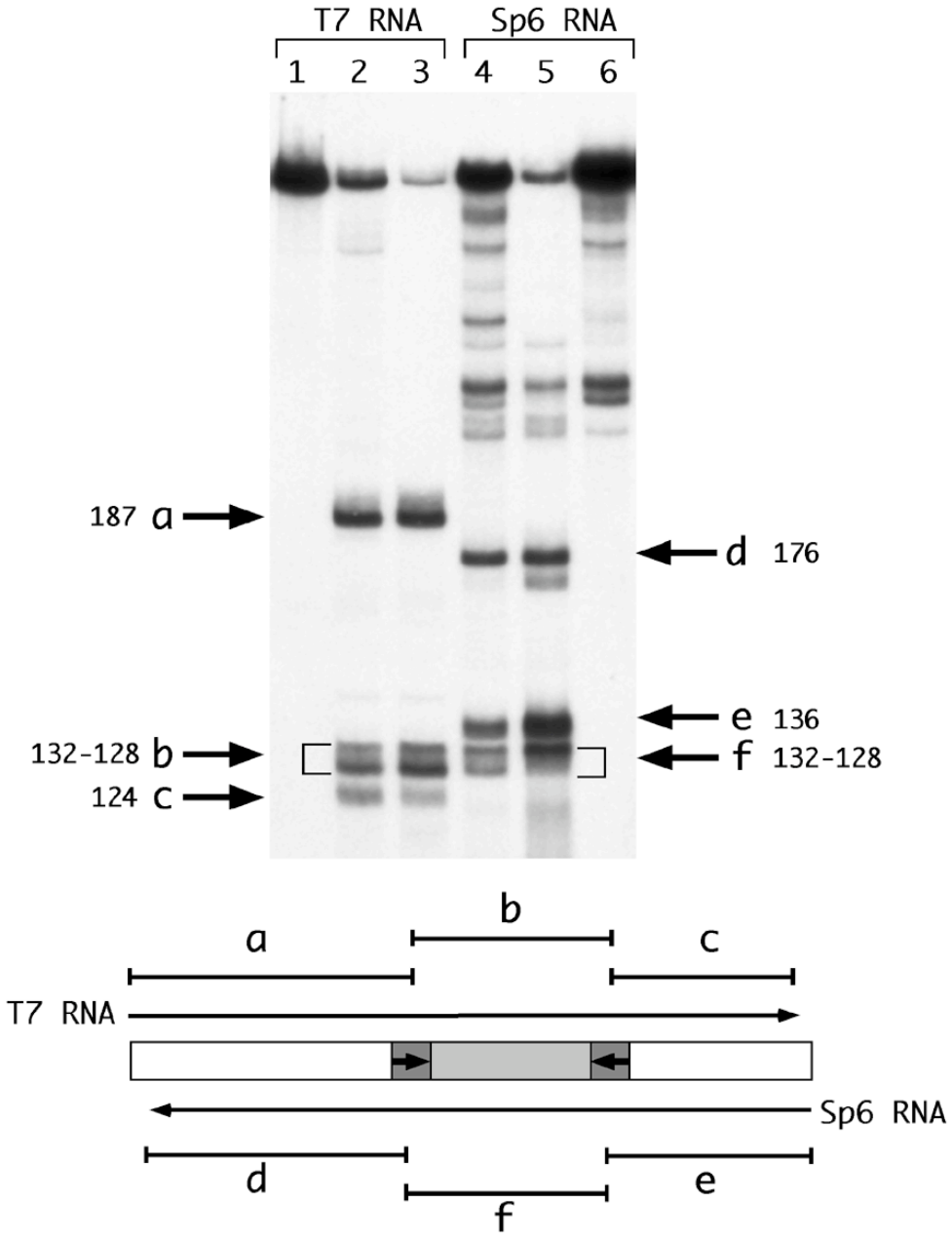


Figura 4. Processamento dei trascritti pGEM-378 *in vitro*. RNA *nemis*+ radioattivi, ottenuti trascrivendo *in vitro* con la T7 o Sp6 polimerasi (canale 1 e canale 6, rispettivamente), sono stati incubati con 0,5 µg (canale 2 e 4) o 1 µg (canale 3 e 5) di estratto cellulare del ceppo wild-type FB1 di *E. coli* per 20' a 37°C. I numeri vicini alle bande a-f indicano la grandezza delle bande in nt, calcolata mediante co-elettroforesi di un marker molecolare. Le frecce nei riquadri del diagramma in basso indicano le TIRs di *nemis*.

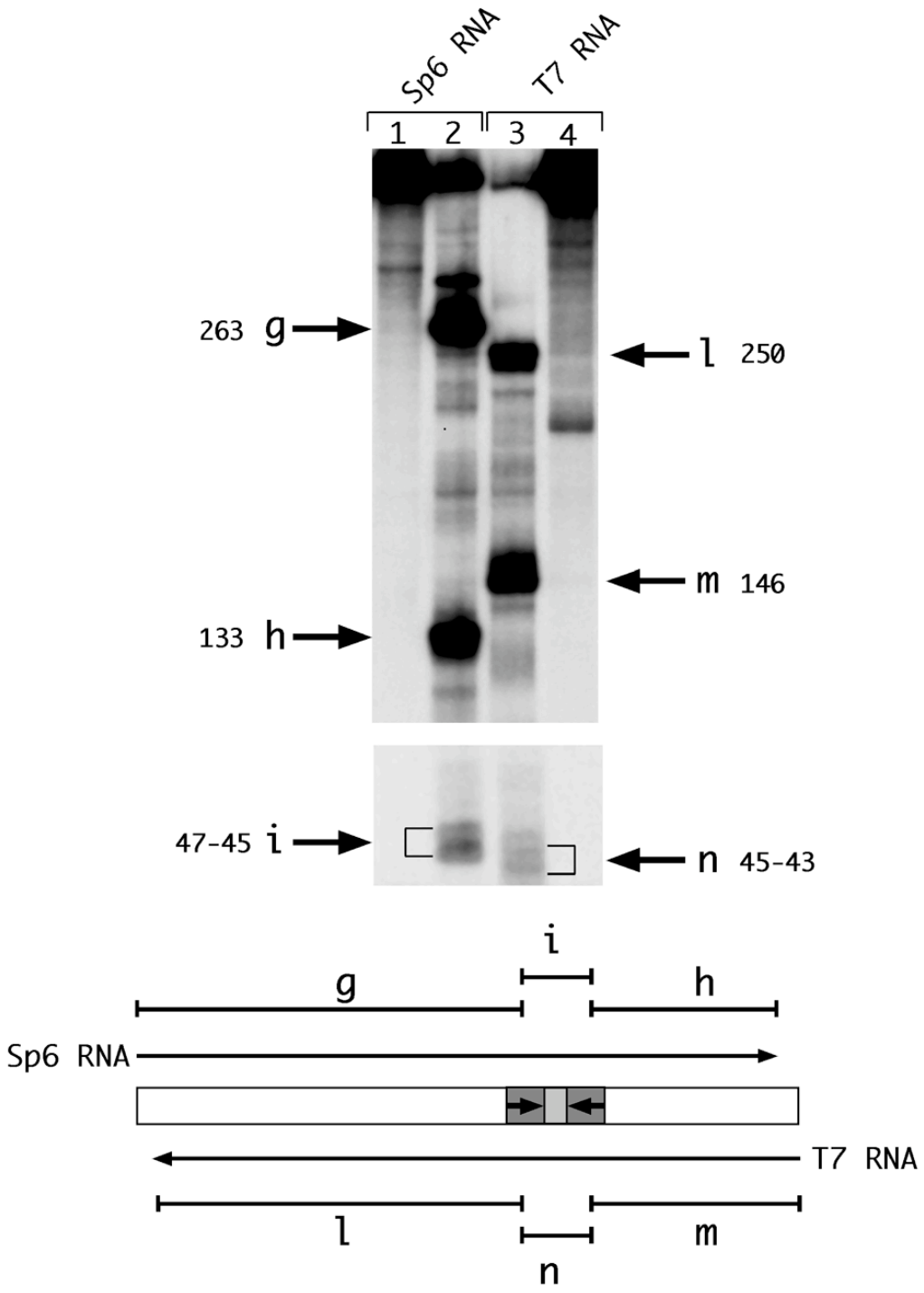


Figura 5. Processamento dei trascritti pGEM-417 *in vitro*. Gli RNA *nemis*+ radioattivi, ottenuti trascrivendo *in vitro* con la T7 o Sp6 polimerasi (canale 1 e canale 4, rispettivamente), sono stati incubati con 0,5 µg (canale 2 e 3) di estratto cellulare del ceppo wild-type FB1 di *E. coli* per 20' a 37°C. La grandezza delle bande g-n è stata calcolata come in Figura 4.

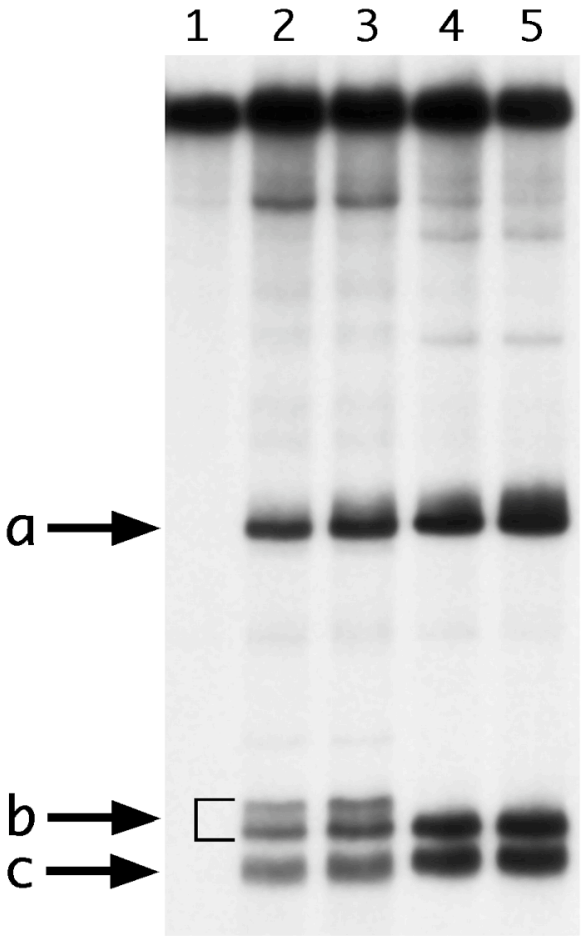


Figura 6. Gli estratti cellulari di *N. lactamica* processano gli RNA *nemis*+. I trascritti pGEM-378 (canale 1) sono incubati con 0,5 µg (canale 2 e 4) o 1 µg (canale 3 e 5) di estratto cellulare del ceppo wild-type FB1 di *E. coli* (canale 2 e 3) o di *N. lactamica* (canale 4 e 5) per 20' a 37°C. La grandezza delle bande a-b è stata è indicata in Figura 4.

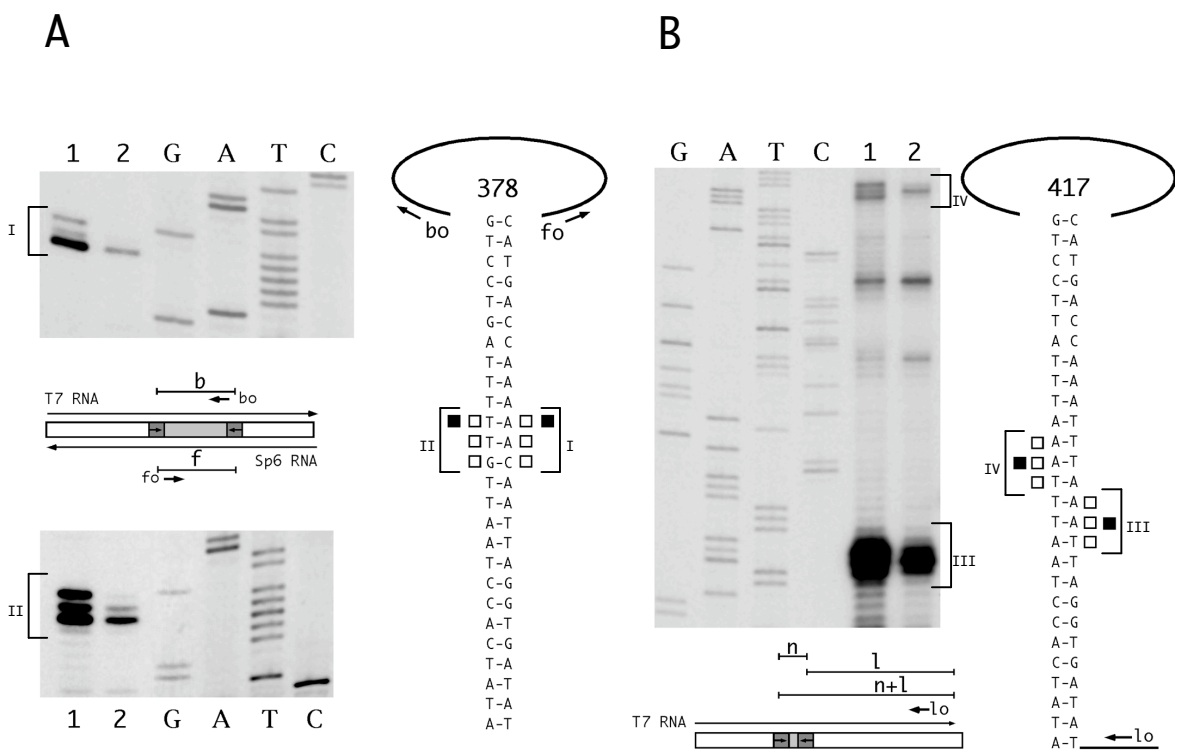


Figura 7. Determinazione dei siti di taglio dei trascritti *nemis*+. A) I trascritti freddi pGEM-378 T7 (pannello superiore) e pGEM-378 Sp6 (pannello inferiore) sono incubati con 0,5 µg di estratto cellulare di *E. coli* (canale 1) o di *N. lactamica* (canale 2) per 20' a 37°C. Le specie processate sono ibridate con gli oligomeri marcati *bo* e *fo*, i quali sono elongati dalla SuperScript II RT. B) I trascritti freddi pGEM-417 T7 sono incubati con 0,5 µg di estratto cellulare di *E. coli* (canale 1) o di *N. lactamica* (canale 2) per 20' a 37°C. Le specie processate sono ibridate con l'oligomero marcato *lo*, il quale è elongato dalla SuperScript II RT. Le bande I-IV indicano i siti di taglio che si trovano dentro TIRs.

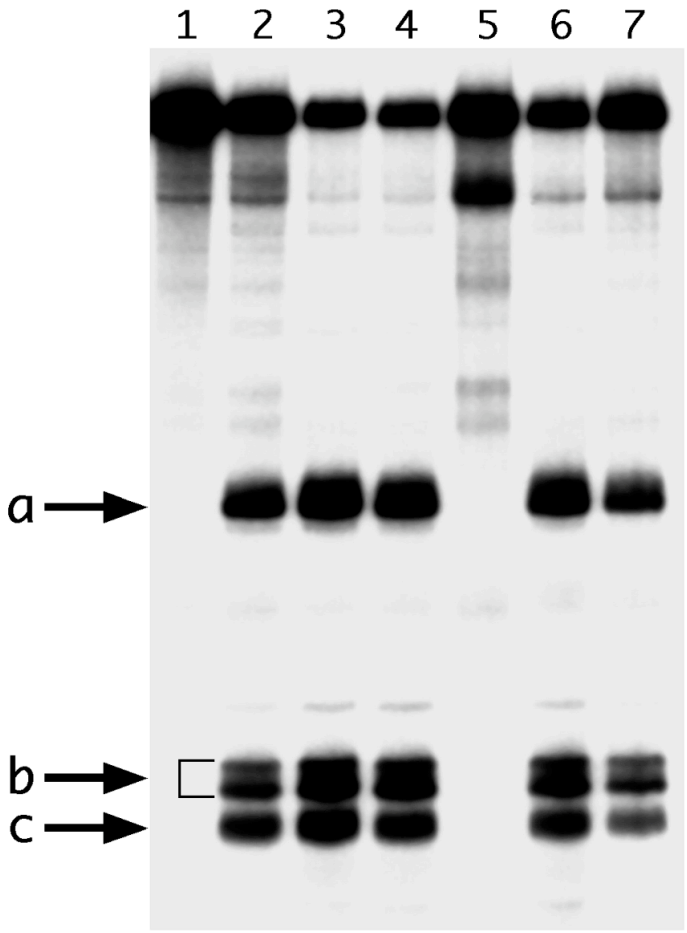


Figura 8. Processamento specifico dei trascritti *nemis*+ da parte della RNasi III. I trascritti pGEM-378 T7 (canale 1) sono incubati per 20' a 37°C con 0,5 µg di estratto cellulare di differenti ceppi di *E. coli:* ceppo wild-type FB1 (canale 2), ceppo SK5003 (canale 3,4), ceppo HT115 (canale 5), ceppo SK5695 (canale 6,7). I ceppi sono stati cresciuti a 32°C. Gli estratti utilizzati nel canale 3 e 7 sono stati preparati dai ceppi SK5003 e SK5695, i quali sono stati cresciuti a 32°C fino in fase logaritmica e poi a 44°C per 45' prima di arrestarne la crescita, al fine di inattivare la RNasi II e la RNasi E, rispettivamente.

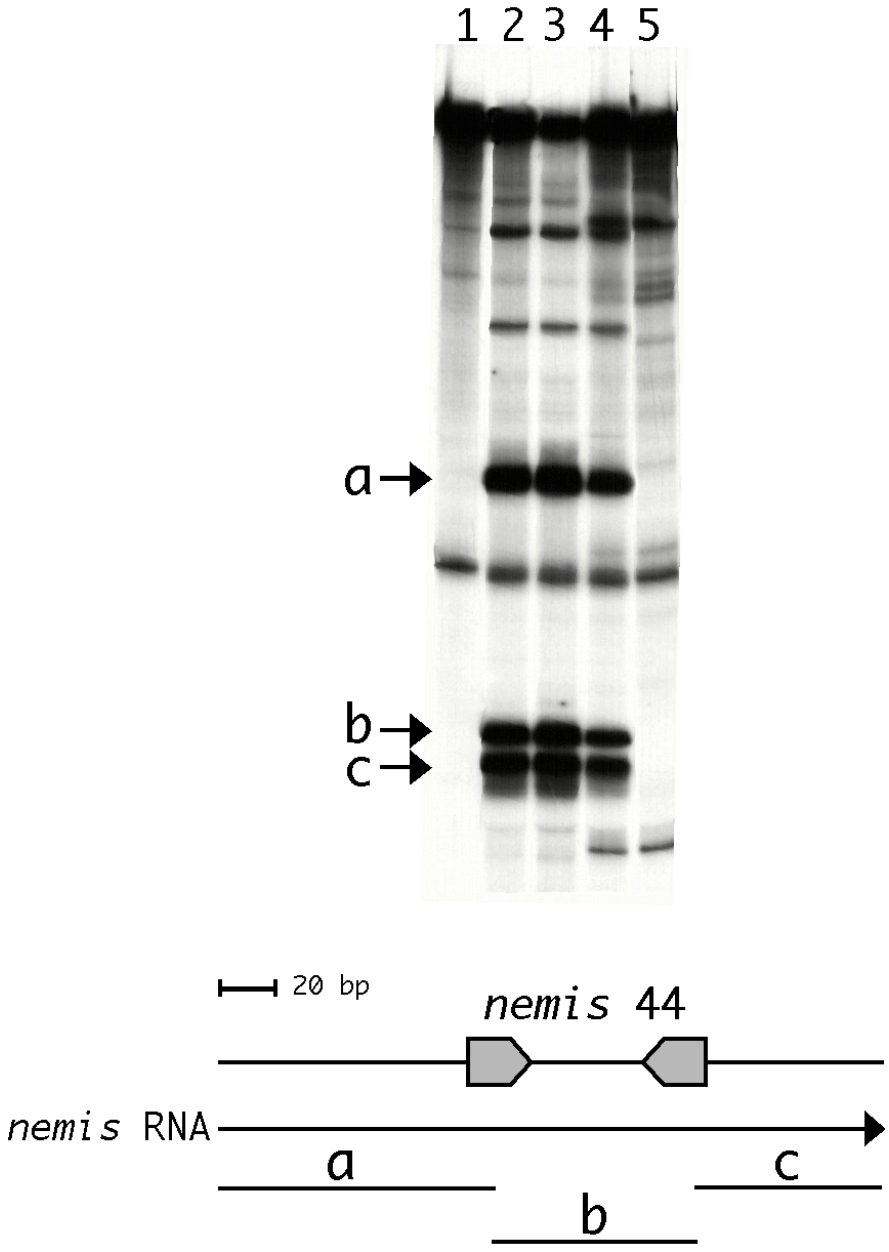


Figura 9. La RNasi III di *N. meningitidis* espressa *in vitro* processa gli RNA *nemis*+. RNA radioattivi che contengono *nemis 44* (canale 1) sono stati incubati per 20' a 37°C con 0,5 µg o 1 µg (canale 2 e 3) di estratto cellulare *N. lactamica* o con la protiena RNasi III di *N. meningitidis* espressa *in vitro* (canale 4). Come controllo i trascritti sono stati incubati con le stesse quantità di estratti di reticolociti usate per esprimere la proteina ricombinante (canale 5).

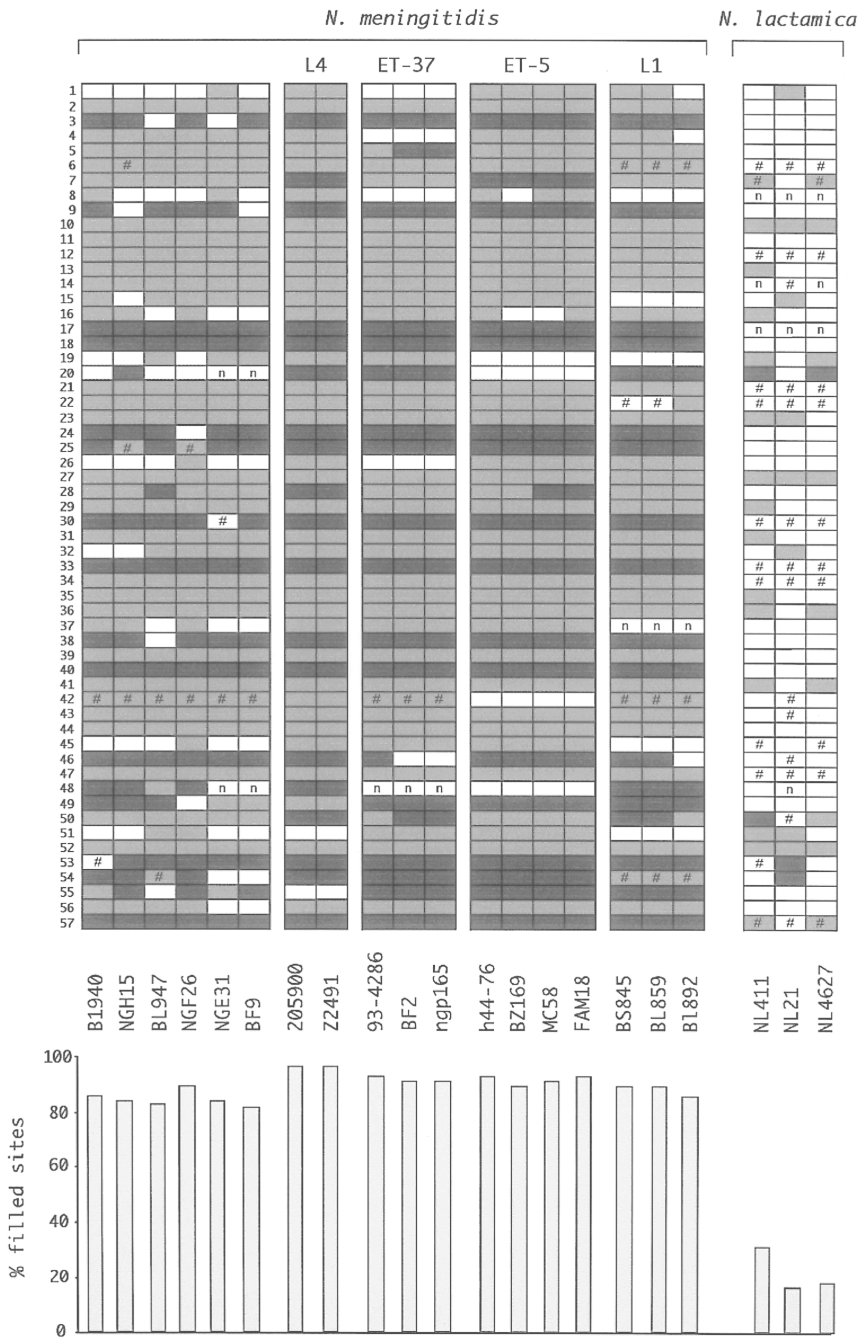


Figura 10. Conservazione degli elementi *nemis* nei genomi delle Neisseriae. I rettangoli vuoti e pieni rappresentano le regioni cromosomali mancanti e contenenti *nemis*, rispettivamente. I rettangoli grigio chiaro e grigio scuro rappresentano i *nemis* normali e riarrangiati, rispettivamente. Le regioni per le quali non si ha amplificazione sono indicate con n. Il simbolo # indica regioni che differiscono dalla grandezza attesa. L'abbondanza relativa delle regioni *nemis*+ in ciascun ceppo è rappresentata nel diagramma in basso.

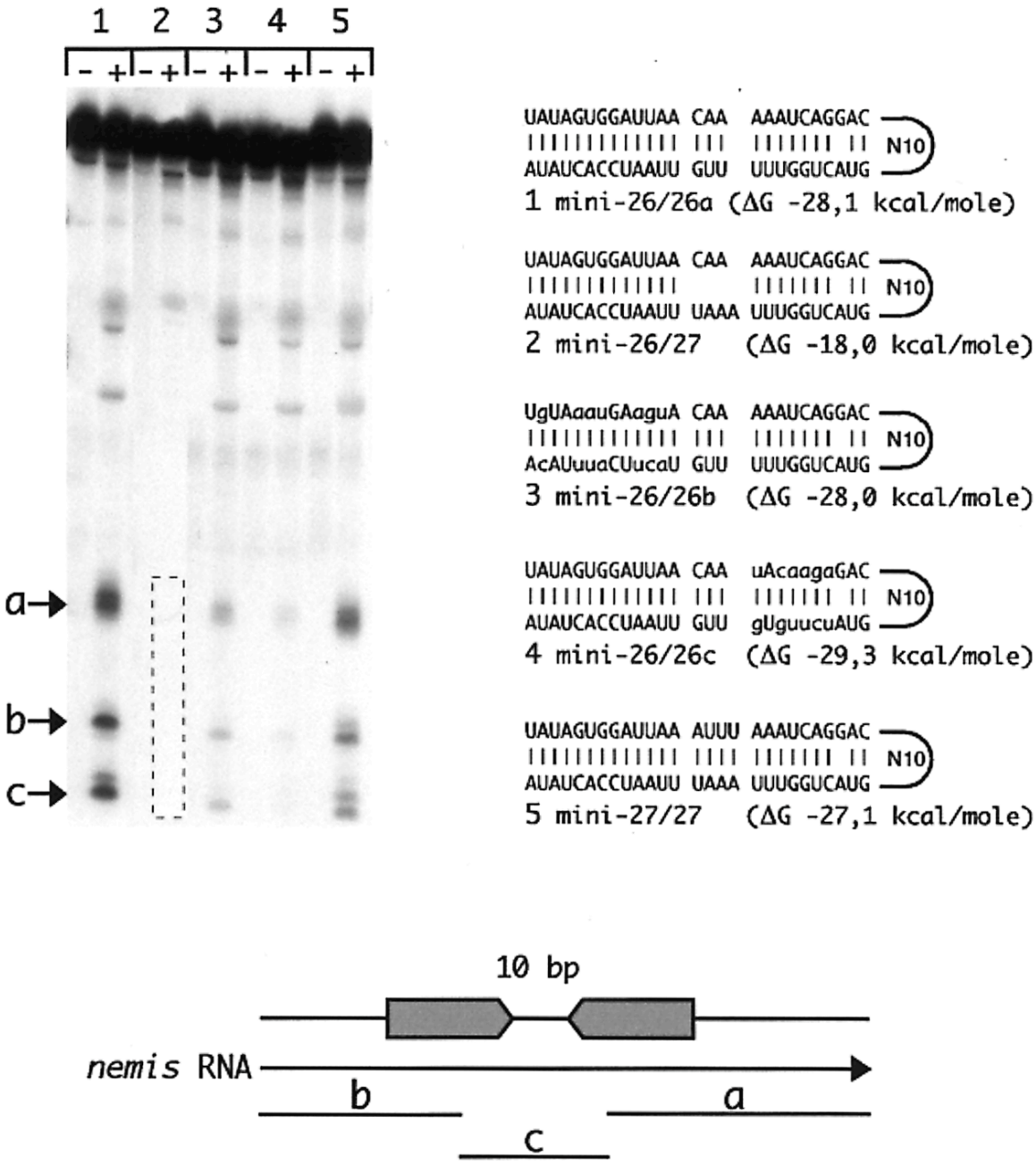


Figura 11. Processamento dei trascritti mini-*nemis*. Gli RNA mini-*nemis*+ radioattivi, ottenuti trascrivendo *in vitro* con la T7 polimerasi, sono stati incubati per 20' a 37°C con e senza la RNasi III di *N. meningitidis* espressa *in vitro* (canale + e -, rispettivamente). La grandezza delle bande a-c è stata calcolata come in Figura 4. Nel pannello di destra sono mostrati il grado di complementarietà e le energie libere delle strutture a forcina formate da ciascun trascritto mini-*nemis*. Le lettere in minuscolo indicano i cambi nucleotidici rispetto alla sequenza *nemis* originale.

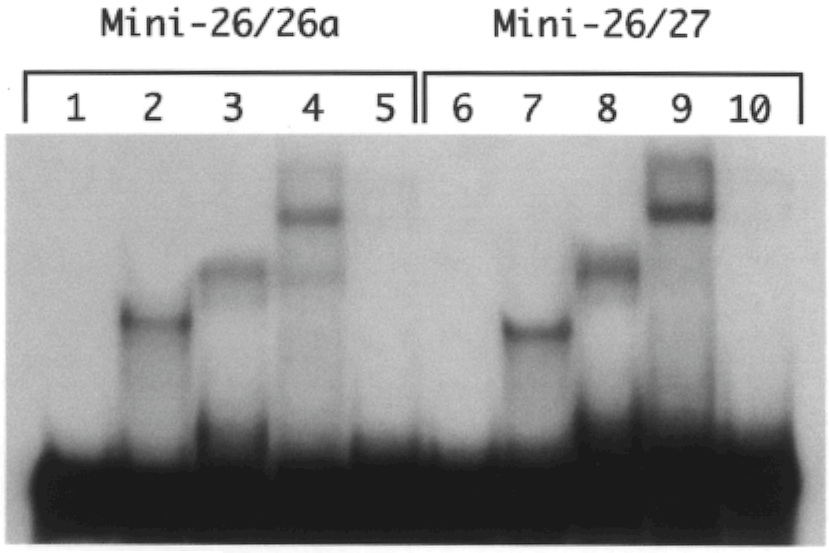


Figura 12. Saggio di *band shift* dei complessi RNA-proteina. Gli RNA radioattivi mini-*nemis* 26/26a e mini-*nemis* 26/27 sono incubati per 30' a 0°C senza (canali 1 e 6) o con la RNasi III di *N. meningitidis* espressa *in vitro* (canali 2 e 7) o con 2 µg di estratto cellulare di *N. lactamica* (canali 3 e 8) o con 2 µg di estratto cellulare del ceppo *rnc*+ di *E. coli* (canali 4 e 9) o con 2 µg di estratto cellulare del ceppo *rnc*- di *E. coli* (canali 5 e 10).

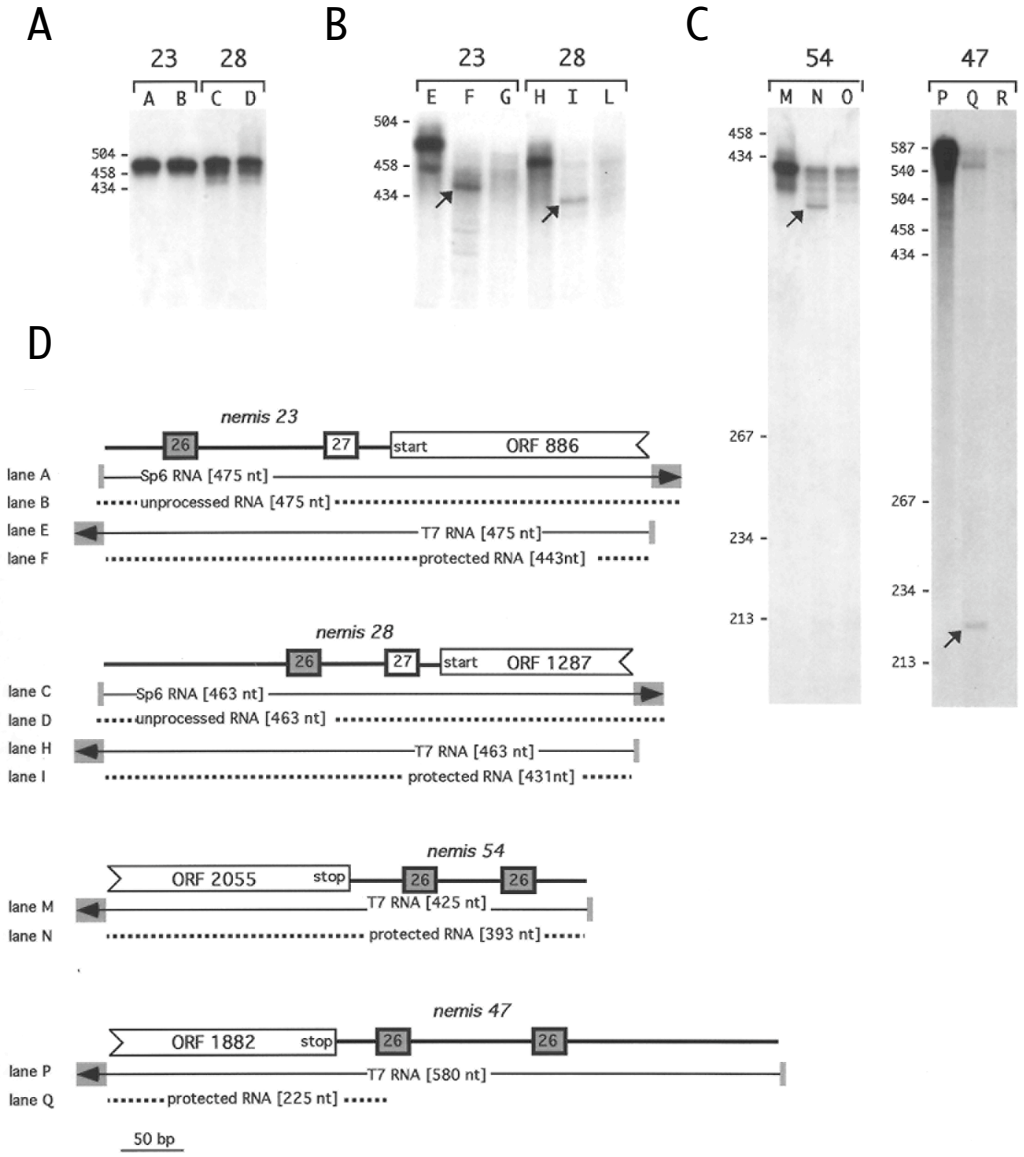


Figura 13. A) Saggio di degradazione *in vitro*. Gli RNA senso radioattivi trascritti *in vitro* dalla Sp6 polimerasi contenenti *nemis* 23 e 28 sono incubati senza (canali A e C) o con 2 µg di estratto cellulare di *N. lactamica* (canali B e D). B) Saggio di *RNase protection*. Gli RNA antisenso radioattivi trascritti *in vitro* dalla T7 polimerasi contenenti *nemis* 23 e 28 (canali E e H) sono ibridati con 5 µg di RNA totali estratti dal ceppo BL859 di *N. meningitidis* (canali F e I) o di RNA di lievito (canali G e L) e digeriti con RNasi T1. C) Saggio di *RNase protection*. Gli RNA antisenso radioattivi trascritti *in vitro* dalla T7 polimerasi contenenti *nemis* 54 e 47 (canali M e P) sono ibridati con 5 µg di RNA totali estratti dal ceppo BL859 di *N. meningitidis* (canali N e Q) o di RNA di lievito (canali O e R) e digeriti con RNasi T1. D) Rappresentazione grafica dei templati contenenti *nemis* 23, 28, 54 e 47 e di parte delle ORFs adiacenti.

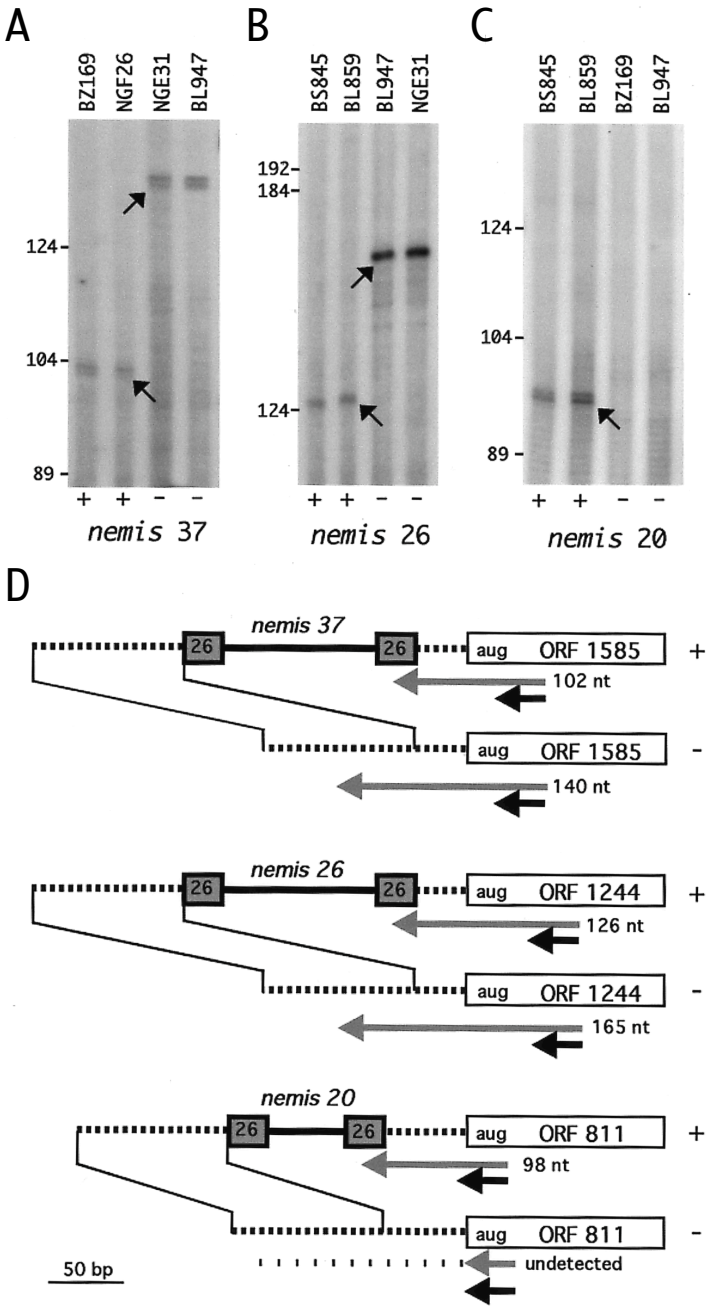


Figura 14. A-C) Saggio di *primer extension*. Oligonucleotidi marcati al 5' con ^{32}P complementari alle ORFs 1585 (A), 1244 (B) e 811 (C) sono ibridati con 5 µg di RNA totali estratti da diversi ceppi di *N. meningitidis* che differiscono per la presenza (+) o assenza (-) di *nemis* 37, 26 e 20 a monte delle ORFs indicate. I principali prodotti di reazione sono indicati dalle frecce. D) Rappresentazione grafica delle regioni analizzate per *primer extension* nei pannelli A-C.

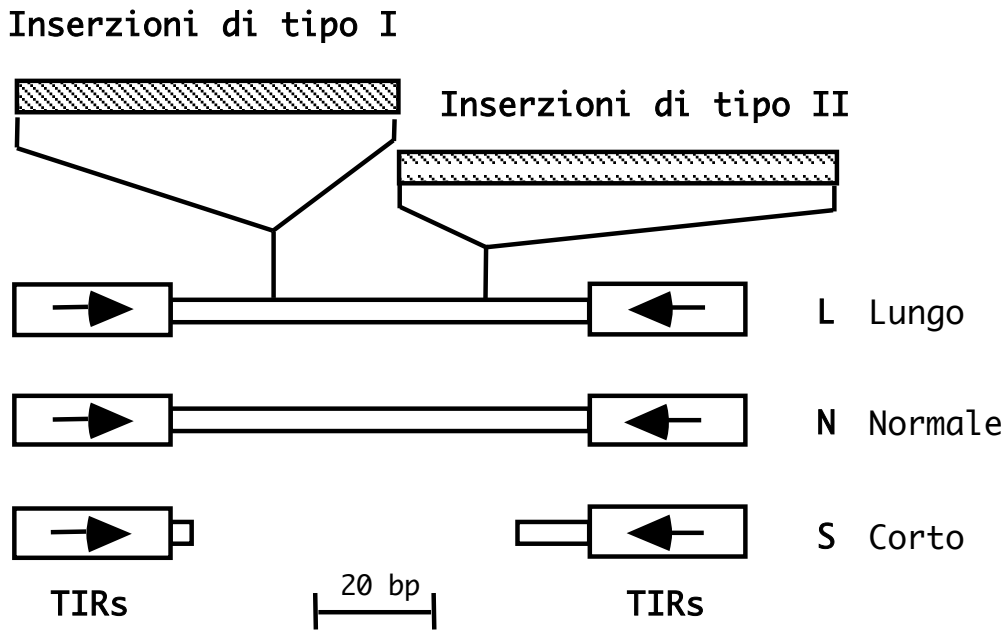


Figura 15. Classi di elementi *eric* individuate nel genoma di *Y. enterocolitica.*

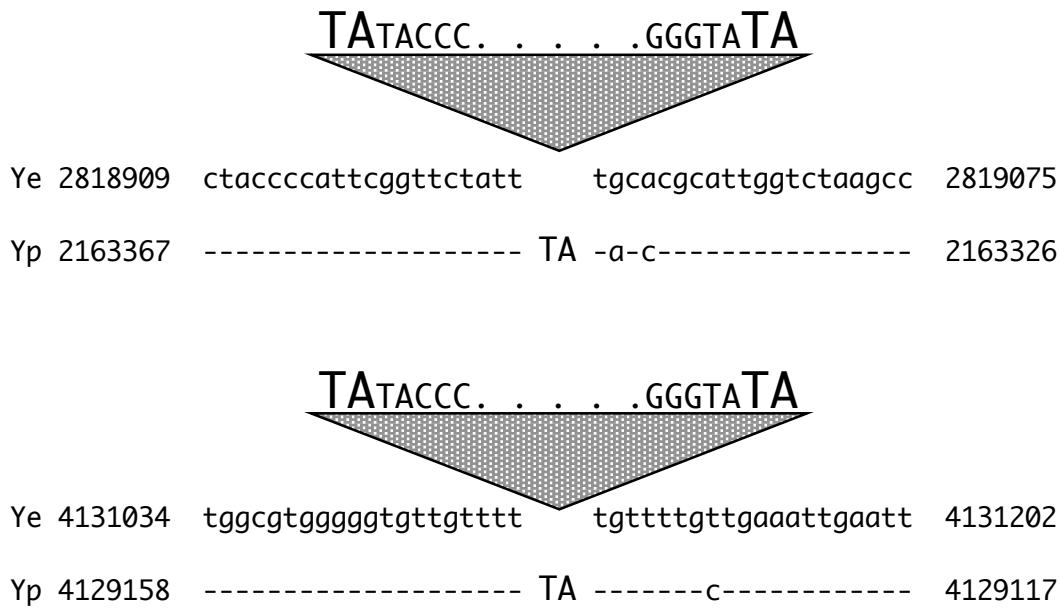


Figura 16. Duplicazione del sito di integrazione di elementi *eric* di *Y. enterocolitica.*

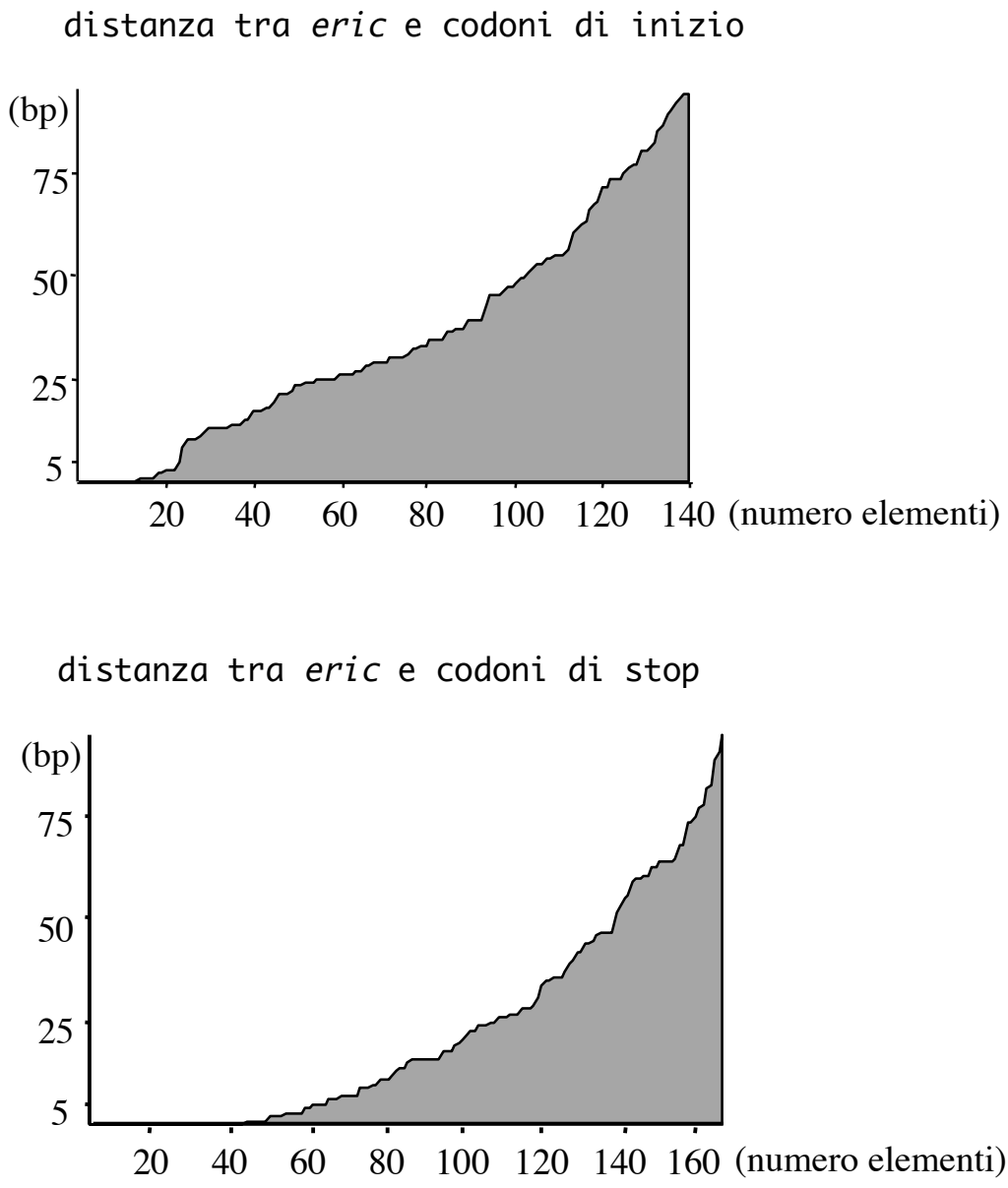


Figura 17. Distanza degli elementi *eric* di *Y. enterocolitica* dalle regioni codificanti adiacenti.

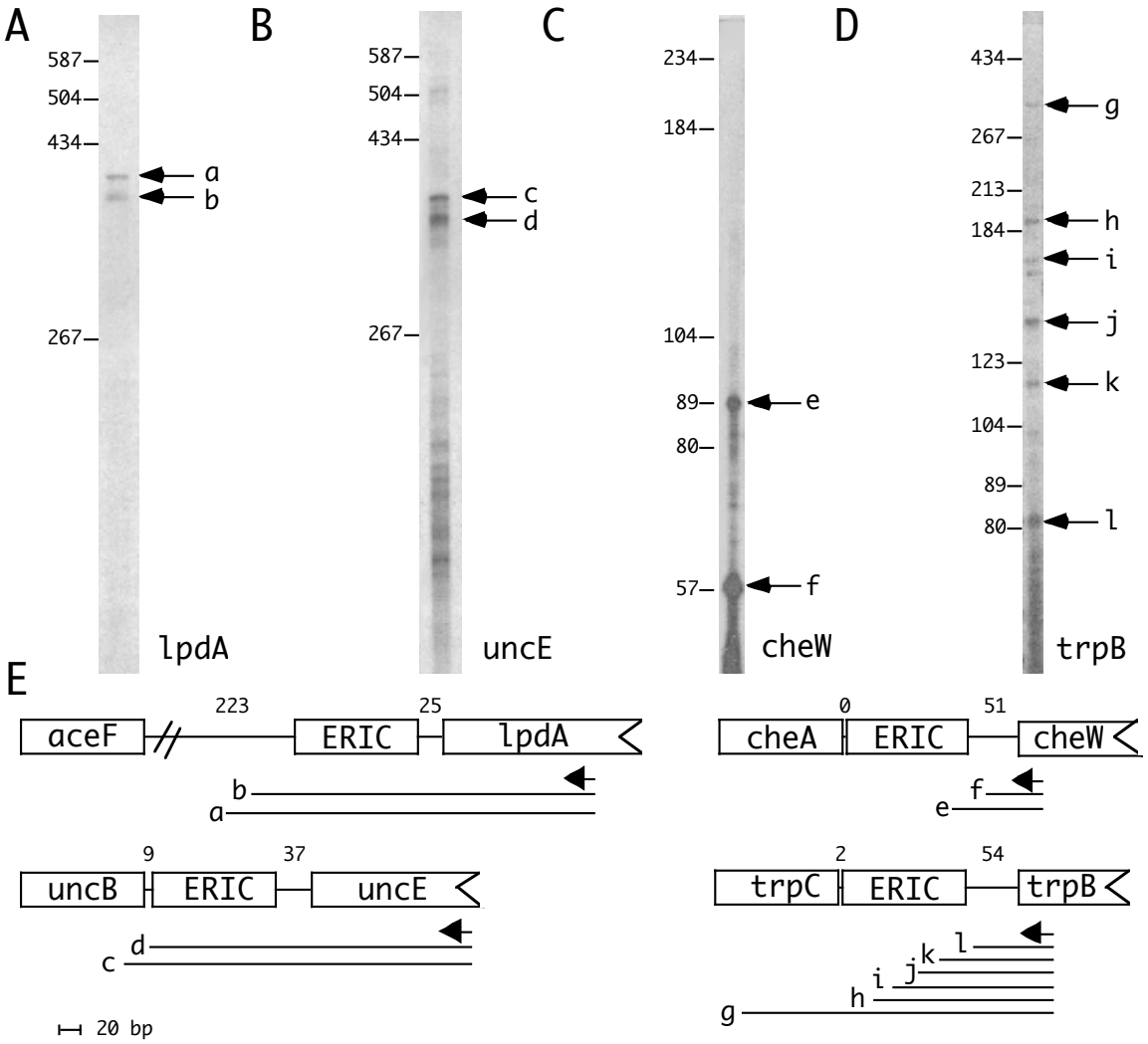


Figura 18. A-D) Saggio di *primer extension* dei trascritti *eric*+. Oligonucleotidi marcati all'estremità 5' con ^{32}P complementari ai geni *lpdA*, *uncE*, *cheW* e *trpB* sono ibridati con 5 µg di RNA totali estratti da Ye-161. I principali prodotti di reazione sono indicati dalle frecce. E) Rappresentazione grafica delle regioni analizzate per *primer extension* nei pannelli A-D. Le frecce rappresentano gli oligonucleotidi utilizzati; i numeri si riferiscono alla distanza in bp che separano gli elementi *eric* dai segnali di inizio e termine della traduzione dei geni adiacenti.

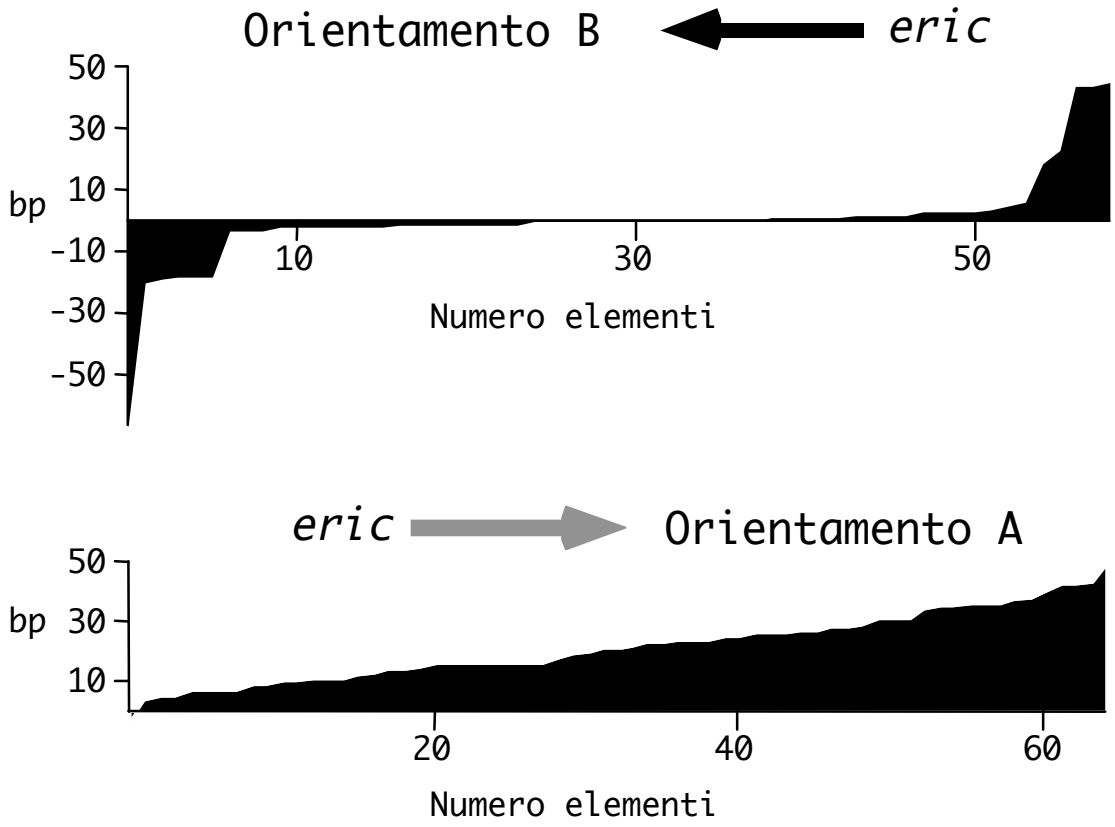


Figura 19. Distanza degli elementi *eric* nell'orientamento A e nell'orientamento B dalle ORFs localizzate a monte.

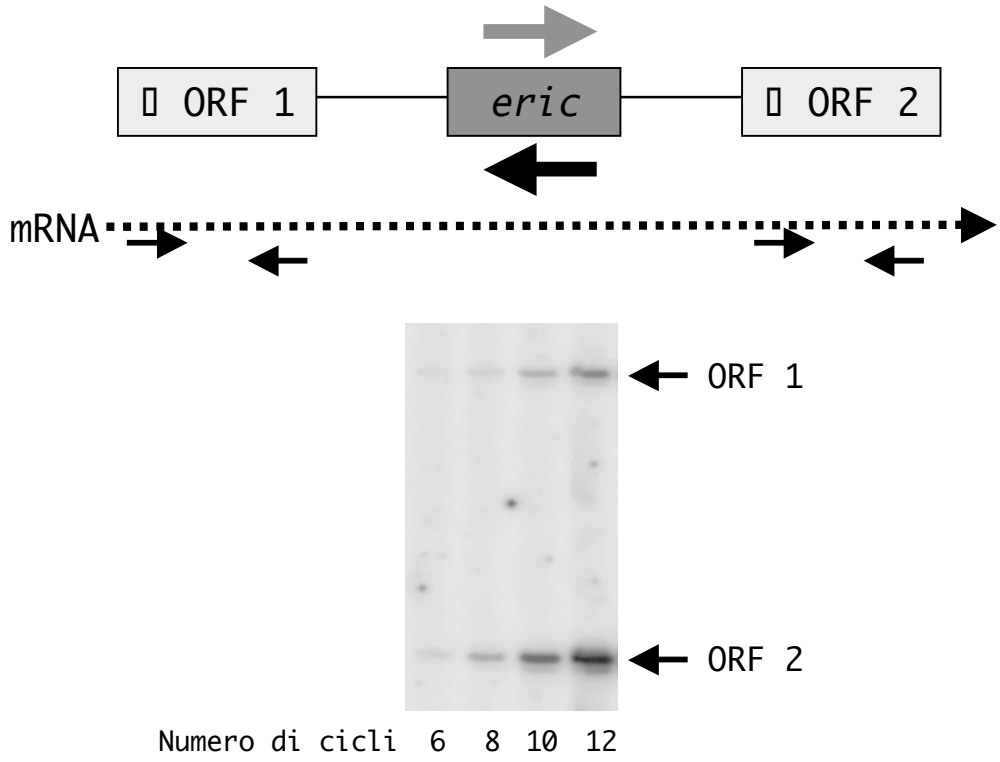


Figura 20. Schema esemplificativo dei saggi di RT-PCR. L' RNA totale di Ye-161 é stato convertito in cDNA utilizzando una miscela di esanucleotidi casuali e successivamenti amplificato mediante PCR usando una specifiche coppie di oligonucleotidi complementari ai geni localizzati a monte e a valle dell'elemento *eric*. In entrambe le coppie, uno dei due oligonucleotidi è stato marcato all'estremità 5' con ^{32}P. I prodotti della reazione sono stati purificati dopo un differente numero di cicli e separati su gel denaturante di poliacrilammide 6% - 8M urea. L'intensità delle bande è stata quantizzata con il PhospoImager.

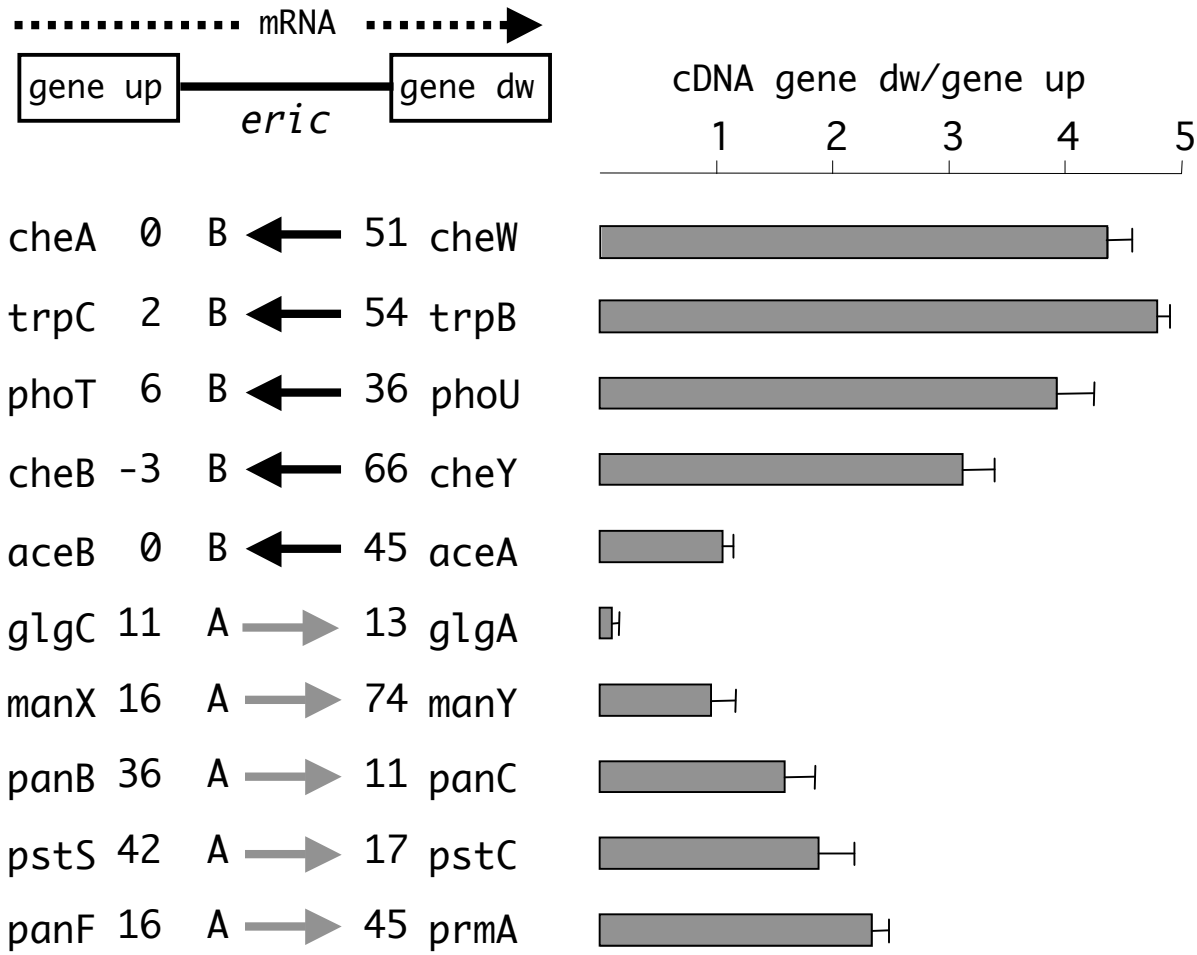


Figura 21. Saggi di RT-PCR. L'RNA totale di Ye-161 é stato convertito in cDNA e successivamenti amplificato mediante PCR. Sono mostrate le coppie di geni analizzati e sono indicate le distanze in bp che separano gli elementi *eric* dai segnali di inizio e termine della traduzione dei geni adiacenti. Per ciascuna coppia di geni è indicato l'orientamento A o B dell'elemento *eric*. La quantità dei trascritti corrispondenti ai geni a valle e a monte è espressa come rapporto. I saggi di RT-PCR sono stati ripetuti da 3 a 4 volte utilizzando due differenti preparazioni di RNA e sono riportate le deviazioni standard.

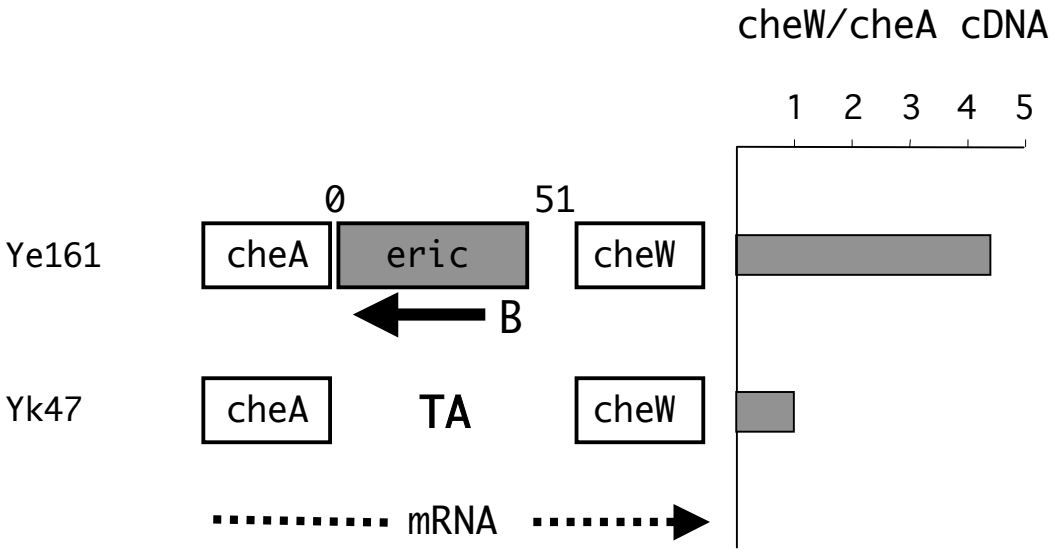


Figura 22. Variazione del rapporto tra i trascritti cheW/cheA associato alla presenza dell'elemento *eric*. Gli RNA totali di Ye-161 e di Yk-47 sono stati convertiti in cDNA e successivamenti amplificati mediante PCR. I numeri si riferiscono alla distanza in bp che separa il codone di stop del gene *cheA* e il codone di start del gene *cheW* dall'elemento *eric*.

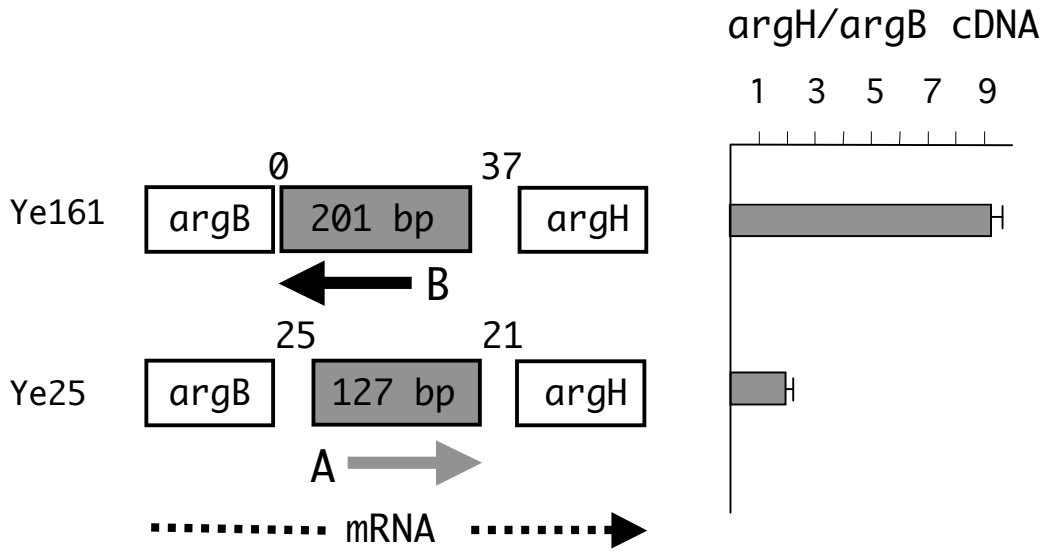


Figura 23. Variazione del rapporto tra i trascritti argH/argB associato alla presenza dell'elemento *eric* nell'orientamento A o nell'orientamento B. Gli RNA totali di Ye-161 e di Ye-25 sono stati convertiti in cDNA e successivamenti amplificati mediante PCR. Sono mostrati la grandezza e l'orientamento dell'elemento *eric* nei due ceppi. I numeri si riferiscono alla distanza in bp che separa il codone di stop del gene *argB* e il codone di start del gene *argH* dall'elemento *eric* nei due ceppi.

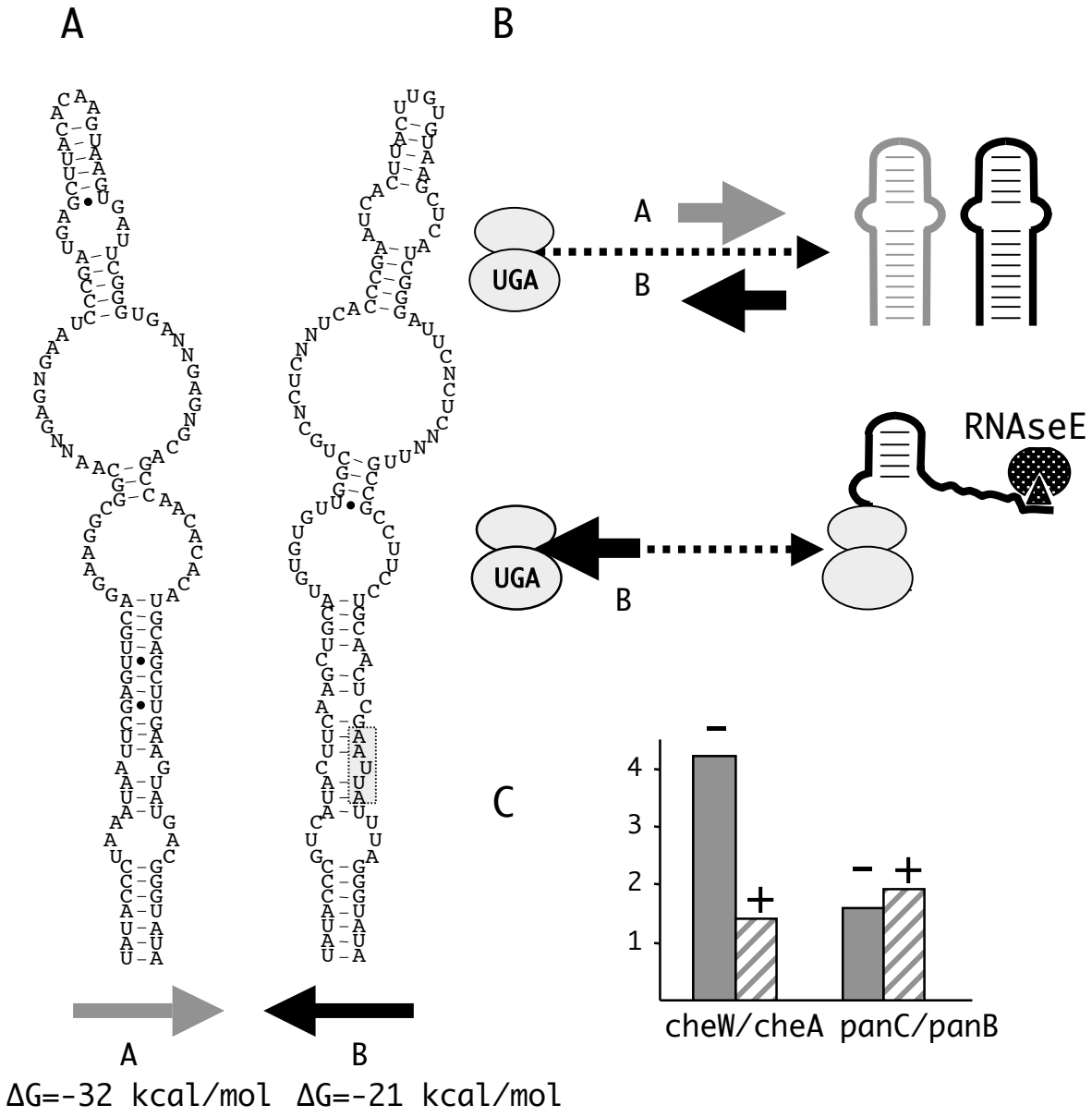


Figura 24. A) Strutture secondarie a forcina degli elementi *eric*. Il riquadro grigio indica il sito bersaglio RAUUW riconosciuto dalla RNase E. B) Quando gli elementi *eric* sono localizzati vicini ai codoni di stop, i ribosomi traducenti non permettono la formazione della struttura a forcina dell'elemento *eric* e il sito bersaglio può essere riconosciuto dalla RNasi E. C) Variazione del rapporto tra i trascritti cheW/cheA (elemento *eric* nell'orientamento B) e del rapporto tra i trascritti panC/panB (elemento *eric* nell'orientamento A) in assenza (-) o in presenza (+) di cloramfenicolo.

LAVORI *IN EXTENSO*

ELSEVIER

Biochimica et Biophysica Acta 1576 (2002) 39–44

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

The abundant class of *nemis* repeats provides RNA substrates for ribonuclease III in *Neisseriae*

Eliana De Gregorio, Chiara Abrescia, M. Stella Carlomagno, Pier Paolo Di Nocera*

Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli Federico II, Via S. Pansini 5, 80131 Naples, Italy

Received 23 October 2001; received in revised form 24 January 2002; accepted 4 February 2002

Abstract

About 2% of the *Neisseria meningitidis* genome is made up by *nemis*, short DNA sequences which feature long terminal inverted repeats (TIRs). Most *nemis* are interspersed with single-copy DNA and are found at close distance from cellular genes. In this work, we demonstrate than RNAs spanning *nemis* of different length and sequence compositions are specifically cleaved at hairpins formed by *nemis* termini by total cellular lysates derived from both *Escherichia coli* and *Neisseria lactamica* strains. The use of cellular extracts from *E. coli* strains impaired in the activity of known ribonucleases let to establish that cleavage at *nemis* TIRs is specifically mediated by the endoribonuclease RNase III. Data set the base for the identification of all of the neisserial genes that are regulated by RNase III because of their physical association with *nemis* DNA. © 2002 Published by Elsevier Science B.V.

Keywords: Repeated DNA family; Terminal inverted repeat; Pathogenic bacteria; RNA hairpin; RNA processing

1. Introduction

Neisseria meningitidis (or meningococcus) and *Neisseria gonorrhoeae* (or gonococcus) are two strictly human pathogens which belong to the same genospecies. Though closely related, the two bacterial species colonize different epithelia and cause notably different diseases. *N. gonorrhoeae* is generally responsible for localized inflammation of the urogenital tract, *N. meningitidis* for meningitis and septicaemia. To rapidly identify potential vaccine candidates, the whole genome sequences of *N. meningitidis* A and B serogroup strains have been recently determined [1,2]. The sequence of a pathogenic *N. gonorrhoeae* strain has also been recently determined (http://dna1.chem.ou.edu/gono.html). Whole genome data provided insights on the unique organization of genetic material in *Neisseriae*. The meningococcus hosts an extraordinary number of different repeated DNA sequences. Of these, many are variously combined in intergenic repeat arrays plausibly involved in recombination processes frequently occurring in *Neisseriae* [1]. Other DNA repeats, as the so-called Correia [3] or *nemis* [4] sequences, are predominantly found as individual units. *Nemis* are miniature DNA insertion sequences (80–160 bp), which feature 26–27 bp long terminal inverted repeats (TIRs). The presence of target site duplications at *nemis* ends, and the identification in sequenced strains of homologous chromosomal regions containing or lacking *nemis* DNA, both indicate that *nemis* are (or have been) mobile genetic elements [4]. Both transposition and recombination events contributed to the genomic spread of this unusual class of DNA repeat family, which is organized in a few major structural subsets and includes, both in A and B serogroup meningococci, ~300 members. Of these, about two-thirds are interspersed with single-copy DNA, and inserted at close distance from coding regions [4]. Both the abundance and the peculiar pattern of interspersion suggest that *nemis* may have regulatory functions. The hypothesis that these repeats may influence gene expression is supported by the finding that in 7/7 *N. meningitidis* genes analyzed, the 5′ termini of the gene transcripts lye within flanking *nemis* sequences [4].

In this work, we provide evidence that *nemis* TIRs fold into hairpin structures, and RNAs encompassing *nemis* sequences are invariably cleaved by cellular lysates from

Abbreviations: bp, base pairs; nt, nucleotide(s); cpm, counts per minute

* Corresponding author. Tel.: +39-081-746-2059; fax: +39-081-770-3285.

E-mail address: dinocera@cds.unina.it (P.P. Di Nocera).

0167-4781/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0167-4781(02)00290-7

either *Escherichia coli* or *Neisseria lactamica* strains at hairpins formed by folding of *nemis* TIRs. Cleavage is specifically mediated by RNase III.

2. Materials and methods

2.1. Bacterial strains and extracts

The *E. coli* strains FB1 (Δ*hisGDCBHAFI-gnd*, *rhaA*), HT115 (*rnc*14::ΔTn10), SK5006 (*thr*, *leu*, pDK39, Cm^r *rnb*500) and its derivatives SK5003 (*pnp*7, *rnb*500) and SK5695 (*rne*1) are described in Ref. [5]. The *N. lactamica* strain 4627 was obtained from the Pasteur Institute. Cellular extracts (S30 and S100) were prepared according to Zubay [6]. *E. coli* strains were grown in LB medium, the *N. lactamica* strain 4627 in GC broth supplemented with 1% Polyvitox (Bio-Merieux).

2.2. Plasmid construction

Plasmids pGEM-378 and pGEM-417 contain *Neisseria* DNA flanking a full-length and an internally rearranged *nemis*, respectively. The pGEM-378 and pGEM-417 inserts are homologous to the intervals 1796373 to 1796751 and 383037 to 383420 of the *N. meningitidis* serogroup B strain MC58 [2], respectively. DNA inserts, obtained by amplifying genomic DNA from the *N. meningitidis* 1055 strain via PCR by using appropriate oligonucleotides as primers, were cloned in the *Sma*I site of the vector pGEM4Z to obtain complementary RNA transcripts, since the vector carries Sp6 and T7 promoters next to the cloning region. Clones were checked by DNA sequencing [7].

2.3. RNA substrates and in vitro cleavage analyses

Uniformly ^{32}P-labeled RNAs were obtained by transcribing in vitro linearized pGEM-378 and pGEM-417 DNAs with either T7 or SP6 RNA polymerase, in reaction mixtures containing ATP, CTP and GTP (500 μM each), UTP (250 μM) and 25 μCi (alpha-^{32}P)-UTP (400 Ci $mmol^{-1}$). Samples aliquots (25,000–30,000 cpm) were incubated with either S30 or S100 bacterial extracts as previously described [5]. Primer extension assays were performed essentially as described [8]. The sequences of the oligomers used as primers in the experiments reported in Fig. 4 are given below:

oligomer **bo** 5′-GCCTTAGCTCAAAGAGAACGATTC-TCTAAG-3′
oligomer **fo** 5′-CAGACAGTACAGACAGATAGTACG-GAACCG-3′
oligomer **lo** 5′-GGACAGGAAGAACACAGCGTTTT-CATCTGA-3′

3. Results

3.1. Transcripts encompassing nemis sequences are cleaved E. coli cellular lysates at specific sites

The DNA fragment cloned in the plasmid pGEM-378 spans the 5′ end region of the *N. meningitidis* gene encoding the orf NM1970 [1] and a flanking, full-length *nemis* element (Fig. 1). The plasmid was used as template to obtain in vitro radiolabeled RNAs of known length directed by either the Sp6 or the T7 polymerase. The incubation of both T7- and Sp6-driven transcripts with *E. coli* S30 cellular extracts resulted in the accumulation of three major classes of RNAs (Fig. 1, lanes 2 to 5). Cleaved RNA species resulted quite stable over incubation times from 20 to 60 min at 37 °C (data not shown). The sizes of the bands marked a to f in Fig. 1 nicely fit with the hypothesis, stemming from RNA extension analyses carried out in vivo [4], that transcripts encompassing *nemis* are cleaved at RNA hairpins formed by *nemis* TIRs. The cleavage pattern was due to soluble activities present in the lysate, since *E. coli*

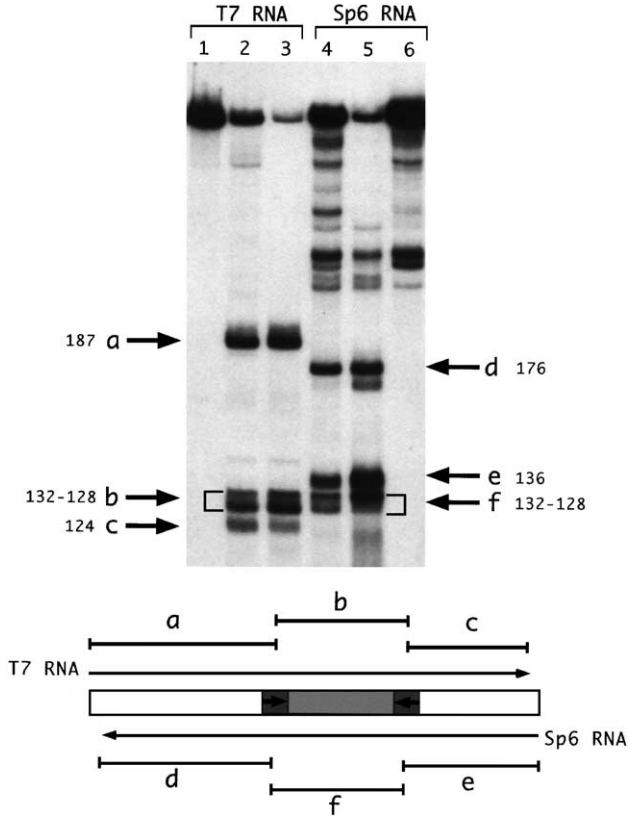


Fig. 1. Specific cleavage occurs at *nemis* RNA. Radiolabeled RNA obtained by transcribing pGEM-378 DNA with either T7 or Sp6 encoded RNA polymerase (lanes 1–3 and 4–6, respectively) were incubated with either 0.5 (lanes 2 and 4) or 1 (lanes 3 and 5) μg of S30 crude cell extracts from the *E. coli* wild-type strain FB1 for 20 min at 37 °C. Reaction products were separated onto a 6% acrylamide-8 M urea gel. Numbers next to bands a to f refer to their length in nt, calculated by co-electrophoresing the reaction products with RNA MW markers not shown in the autoradiogram. Boxed arrows in the diagram at the bottom denote *nemis* TIRs.

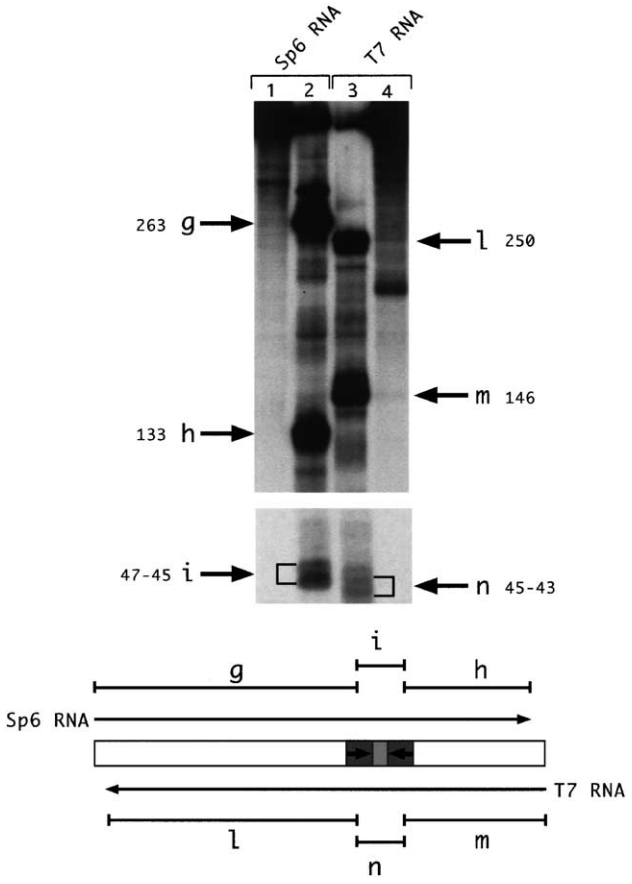


Fig. 2. Cleavage of pGEM-417 transcripts. Radiolabeled pGEM-417 RNA (lanes 1 and 4) were incubated with 0.5 μg of *E. coli* S100 extracts (lanes 2 and 3). The size of bands g to n were determined as in Fig. 1.

S100 cellular extracts produced the same RNA species shown in Fig. 1 (data not shown). Subsequent experiments were therefore carried out with S100 cellular fractions.

The full-length *nemis* cloned in pGEM-378 features 26-bp-long TIRs. Several *nemis* are rearranged because of the loss of a 50-bp-long internal segment. Moreover, elements featuring 27 bp long TIRs, which partly differ in sequence from 26 bp long TIRs, also occur [4]. Different members of the nemis family could generate alternative RNA structure, and RNA cleavage could be influenced by either the sequence content of *nemis* and/or the distance between TIRs. To verify the issue, transcripts from pGEM-417, a plasmid which contains an internally rearranged, 81 bp long *nemis* featuring 27 bp long TIRs, were also challenged with the *E. coli* cellular extracts. Bands marked g to i and l to n in Fig. 2 have the size expected for RNA moieties produced by cleavage of the pGEM-417 substrate at *nemis* TIRs. We also set up competition experiments in which radiolabeled pGEM-417 transcripts were challenged with cellular lysates in the presence of two to five hundred-fold excesses of cold RNA. Cleavage of radiolabeled pGEM-417 transcripts was competed at comparable levels by cold excess of pGEM-417 and pGEM-378 transcripts (data not shown). On the whole, data indicate that transcripts encompassing *nemis* elements, which differ either in size or sequence content, are processed, at least in vitro, with the same efficiency.

3.2. A cellular lysate derived from N. lactamica cleaves nemis RNA

Challenging pGEM-378 transcripts with cellular extracts from a wild type strain of the apathogenic *N. lactamica* species gave essentially the same cleavage pattern produced by the *E. coli* extracts (Fig. 3). The size of the RNA species measured in the experiments shown in Figs. 1–3 is not definitive proof that cleavages occur at *nemis* inverted repeats. Such hypothesis was confirmed by primer extension analyses shown in Fig. 4. The experiments were performed by incubating first cold *nemis* RNAs (either pGEM-378 or pGEM-417) with either *E. coli* or *N. lactamica* cellular lysates. Specific oligonucleotides were subsequently annealed to the in vitro processed b (**bo** oligomer) f (**fo** oligomer), l and n+l (**lo** oligomer) RNA species (see also Fig. 1), and extended with the reverse transcriptase to map cleavage sites. Products of extensions on pGEM-378 transcripts, driven by either the T7 (bands I, panel A, top) or the Sp6 promoter (bands II, panel A, bottom), were obtained in separate experiments. In panel B, bands III and IV denote

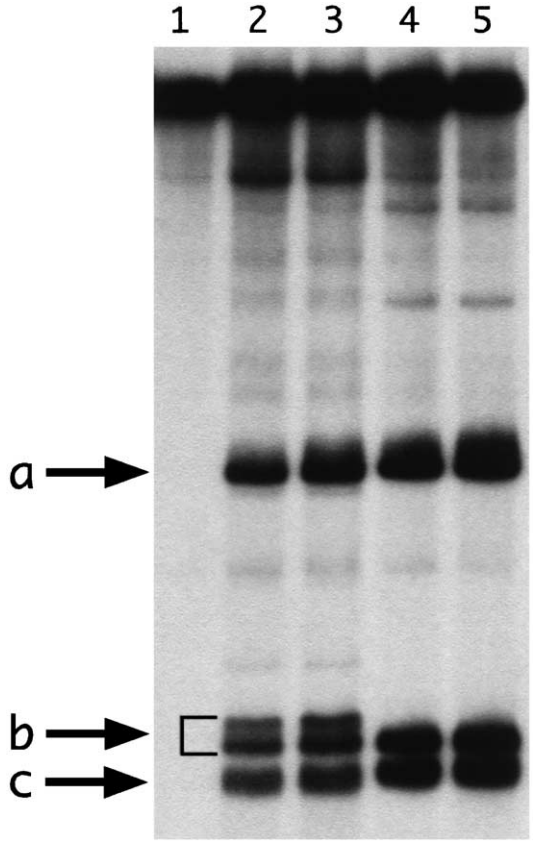


Fig. 3. *N. lactamica* lysates cleave *nemis* RNA. pGEM-378 transcripts (lane 1) were incubated with either 0.5 (lanes 2 and 4) or 1 μg (lanes 3 and 5) of crude S100 extracts from either *E. coli* (lanes 2 and 3) or *N. lactamica* (lanes 4 and 5).

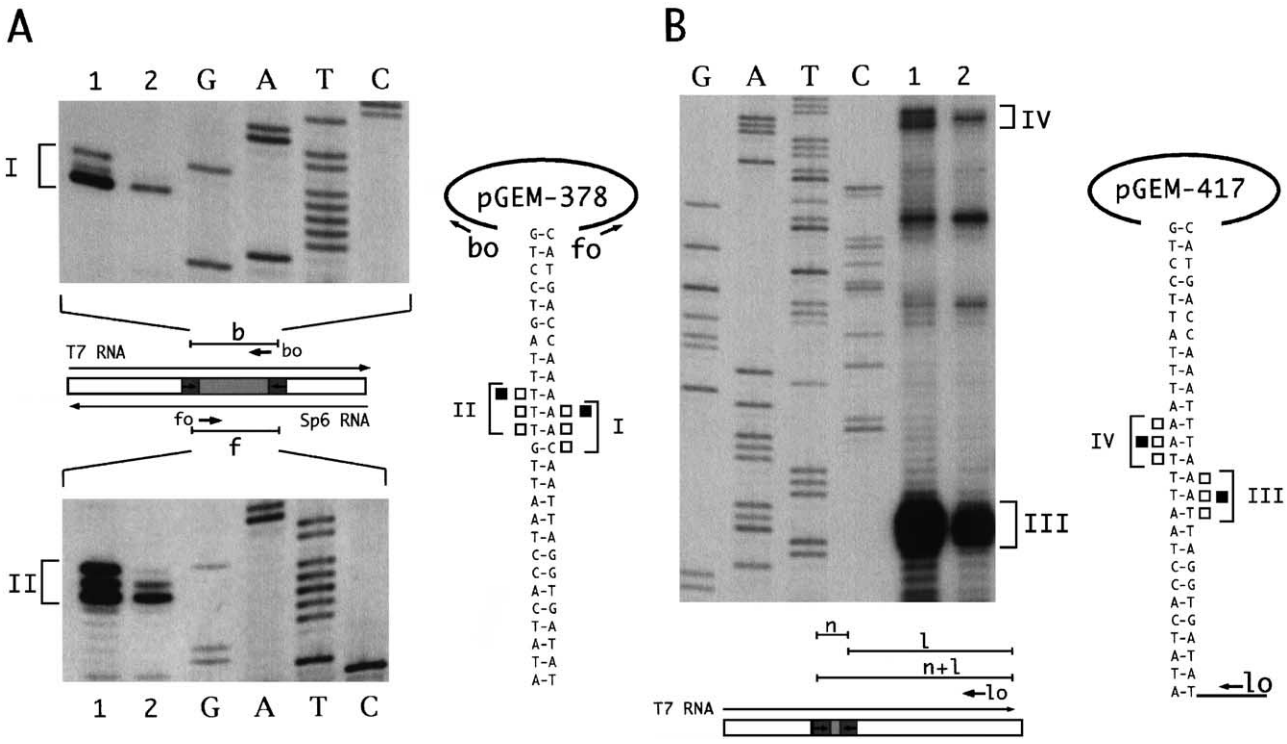


Fig. 4. Cleavage sites at *nemis* hairpins. (A) Cold RNA substrates, obtained by transcribing pGEM-378 with either T7 (upper panel) or Sp6 polymerase (lower panel), were challenged with either *E. coli* (lanes 1) or *N. lactamica* (lanes 2) lysates. Processed RNA species were subsequently mixed with the ^{32}P-5′-end-labeled DNA oligomers **bo** and **fo**. Annealed primer moieties were elongated, in the presence of deoxynucleosides triphosphates, by the SuperScript II reverse transcriptase. Reaction products were separated onto a 8% acrylamide-8 M urea gel. (B) pGEM-417 transcripts directed by the T7 polymerase were challenged with either *E. coli* or *N. lactamica* lysates (lanes 1 and 2, respectively). Samples were subsequently mixed with the ^{32}P-5′-end-labeled DNA oligomer **lo**, and processed as in panel A. Sequencing ladders of the plasmids pGEM-378 and pGEM-417 were obtained by the dideoxy chain termination method by using the **bo**, **fo** and **lo** oligomers as primers. Bands I to IV denote cleavage sites within *nemis* TIRs. Empty and filled squares mark cleavage sites mapped with *E. coli* and *N. lactamica* lysates, respectively.

the elongation products obtained with processed pGEM-417 transcripts driven by the T7 promoter. The additional bands in lanes 1 and 2 of panel B of Fig. 4 may denote either premature stops of the reverse transcriptase or elongation products of minor processed species of pGEM-417 RNA. Slight differences in the cleavage specificity of the *E. coli* and *N. lactamica* processing activities are highlighted in Fig. 4.

3.3. Nemis RNA hairpins are targeted by RNase III

Radiolabeled pGEM-378 transcripts were incubated with S100 cellular extracts derived from *E. coli* strains harboring mutant alleles for specific ribonucleases. Only extracts from HT115, a strain carrying a *rnc* allele encoding an inactive endoribonuclease RNase III [9], were totally unable to cleave the RNA substrate (Fig. 5, lane 5). The prominent band immediately below the input RNA band in lane 5 of Fig. 5 likely corresponds to uncleaved substrate moieties in which the formation of secondary structures caused a block to 3′ exonuclease activities. This conclusion is supported by the finding that the same RNA species is relatively less abundant in samples incubated with cellular extracts from the strain SK5003, in which a mutation in the *pnp* gene inactivates the 3′ exonuclease activity of the phosphonucleotide phosphorylase (Fig. 5, lanes 3 and 4). It is of interest to note that the mRNA-specific RNase E has no role in the process of cleaving *nemis* RNA (Fig. 5, lanes 6 and 7).

4. Discussion

Bacterial cells possess several endoribonuclease activities, which function both in the processing of stable RNAs and the mRNA decay [10–14]. Intergenic sequences may play an important role in the control of RNA enzymatic decay. Both in *E. coli* and *S. typhimurium* REPs, short DNA repeats inserted within operons, protect mRNAs from 3′–5′ exonuclease decay, by forming stem-loop structures able to suppress the action of 3′ endoribonucleases [15–17]. *Nemis*, an abundant class of DNA repeats uniquely found in *Neisseriae*, represent an additional example of intergenic sequences which may influence the expression of neighbouring genes by acting at the RNA level. We recently reported that the 5′ end termini of several *N. meningitidis* transcripts mapped within *nemis* TIRs [4]. Here, we showed

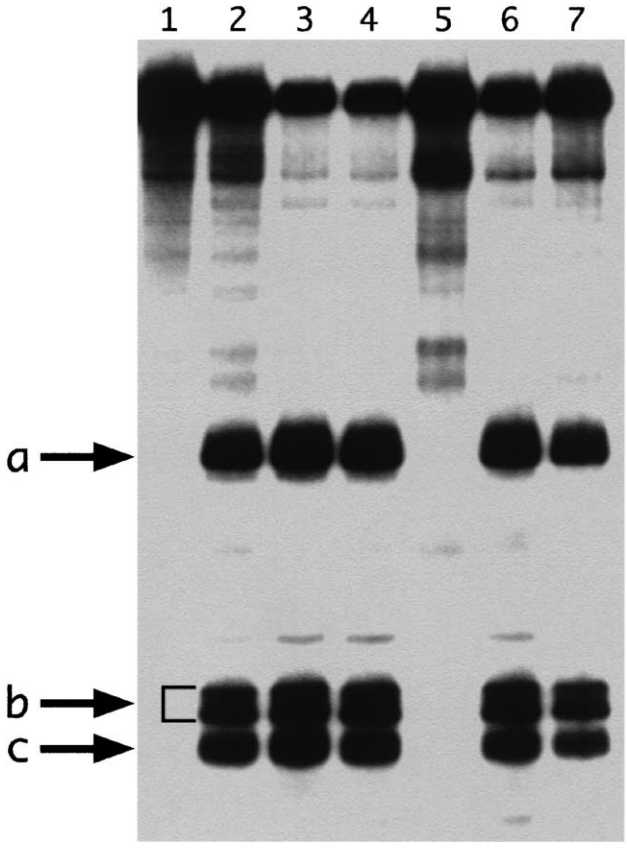


Fig. 5. The endonucleolytic activity responsible for *nemis* RNA cleavage is RNase III. PGEM-378 RNA transcribed by the T7 promoter was incubated for 20 min at 37 °C either alone (lane 1) or with 0.5 μg of S100 cellular lysates from the *E. coli* strains FB1 (lane 2), SK5003 (lanes 3, 4) HT115 (lane 5), SK5695 (lanes 6, 7) grown at 32 °C. Extracts in lanes 4 and 7 derived from SK5003 and SK5695 cells grown to early logarithmic phase at 32 °C and shifted to 44 °C for 45 min before harvesting, to inactivate RNases II and E encoded by the *rne*1 and the *rneB*500 alleles, respectively. Reaction products were separated onto a 6% acrylamide-8 M urea gel.

that cellular lysates from either *E. coli* or *N. lactamica* strains specifically cleave transcripts encompassing *nemis* sequences at hairpins formed by complementary *nemis* TIRs (Figs. 1–3). Double helical RNA structures formed by the folding of *nemis* termini are cleaved at either site, the location of scissile bonds varying in hairpins of different sequence composition (Fig. 4). Cleaved RNA species are rather stable. Plausibly, processed RNA moieties are still kept by hydrogen bonds in robust secondary structures, and hence are protected by exonucleolytic attack. By assaying extracts from *E. coli* strains carrying mutant alleles of genes encoding specific ribonucleases, we settled that the processing activity corresponds to RNase III (Fig. 5).

The *N. meningitidis* genome encodes a 239-aminoacid-long protein homologous to ribonucleases III [1,2] exhibiting 49% of homology (66.7% similarity) to the *E. coli rnc* gene product. RNases III include two functional modules, one responsible for recognition and binding to the target (RIBOc domain), the other for endonucleolytic activity (Double Stranded Recognition Motif or DSRM; see Ref. [18]). By using the SMART program [19], we found that both motifs are conserved in the *Neisseria rnc* gene product, the RIBOc domain starting at position 26 and ending at position 154, the DSRM domain starting at position 161 and ending at position 229 (not shown).

In *E. coli*, bulk mRNA is processed by RNase E [11,14]. The preeminent role of RNase III is to cleave the primary transcript of the *rrn* operons, to provide the immediate precursors to the mature 16S and 23S ribosomal RNA species. Nevertheless, RNase III is known to cleave a few mRNAs, either increasing the stability or accelerating the decay of specific phage and cellular transcripts [10,13,14,20]. Similarly, cleavage of mRNAs reading through *nemis* may provide a mechanism controlling at the post-transcriptional level the half-life of a large number of specific mRNAs in *Neisseriae*. To our knowledge, this is the first evidence that members of an abundant repetitive DNA family provide a series of RNA substrates which are recognized and cleaved by RNase III in prokaryotes. Both the abundance and the genomic distribution of *nemis*, which are preferentially inserted near to cellular genes, suggest that RNase III may act as global regulator in *Neisseriae*. The hypothesis that cleavage of *nemis* RNA plays an important regulatory role is supported by in silico surveys indicating that several transcriptional units are conserved in both pathogenic *N. gonorrhoeae* and *N. meningitidis* strains, but contain *nemis* sequences in the meningococcal strains only [[4]; P.P. Di Nocera, unpublished results]. Thus, intriguingly, post-transcriptional processes involving *nemis* sequences may contribute to the differential expression of homologous genes in different species of the genus *Neisseria*. It will be of interest to ascertain whether specific *nemis* RNA hairpins, either per se or because of the RNA context in which they are embedded, may be refractory to cleavage by RNase III, as the enzyme domain endowed with endonucleolytic activity is not activated in all dsRNA–RNase III interactions [18]. The presence/absence of *nemis* at specific chromosomal locations might as well account for differences in the intracellular levels of specific mRNAs among pathogenic and non-pathogenic meningococcal strains, and future investigations are aimed at clarifying this important issue.

Acknowledgements

This work has been supported by grants from MURST-PRIN and MURST-CNR Biotechnology L95/95 programs.

References

[1] J. Parkhill, M. Achtman, K.D. James, S.D. Bentley, C. Churcher, S.R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R.M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M.A. Quail, M.-A. Rajandream, K.M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B.G. Spratt, B.G. Barrell, Complete DNA sequence of a serogroup

A strain of *Neisseria meningitidis* Z2491, Nature 404 (2000) 502–506.

[2] H. Tettelin, N.J. Saunders, J. Heidelberg, A.C. Jeffries, K.E. Nelson, J.A. Eisen, K.A. Ketchum, D.W. Hood, J.F. Peden, R.J. Dodson, W.C. Nelson, M.L. Gwinn, R. DeBoy, J.D. Peterson, E.K. Hickey, D.H. Haft, S.L. Salzberg, O. White, R.D. Fleischmann, B.A. Dougherty, T. Mason, A. Ciecko, D.S. Parksey, E. Blair, H. Cittone, E.B. Clark, M.D. Cotton, T.R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H.O. Smith, C.M. Fraser, E.R. Moxon, R. Rappuoli, J.C. Venter, Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58, Science 287 (2000) 1809–1815.

[3] F.F. Correia, S. Inouye, M. Inouye, A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA, J. Bacteriol. 167 (1986) 1009–1015.

[4] M. Mazzone, E. De Gregorio, A. Lavitola, C. Pagliarulo, P. Alifano, P.P. Di Nocera, Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*, Gene 278 (2001) 211–222.

[5] P. Alifano, F. Rivellini, C. Piscitelli, C.M. Arraiano, C.B. Bruni, M.S. Carlomagno, Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA, Genes Dev. 8 (1994) 3021–3031.

[6] G. Zubay, In vitro synthesis of protein in microbial systems, Annu. Rev. Genet. 7 (1973) 267–287.

[7] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

[8] E. De Gregorio, L. Chiariotti, P.P. Di Nocera, The overlap of Inr and TATA elements sets the use of alternative transcriptional start sites in the mouse galectin-1 gene promoter, Gene 268 (2001) 215–223.

[9] H.E. Takiff, S.M. Chen, D.L. Court, Genetic analysis of the *rnc* operon of *Escherichia coli*, J. Bacteriol. 171 (1989) 2581–2590.

[10] D.L. Court, RNA processing and degradation by RNase III, in: G.J. Belasco, G. Brawerman (Eds.), Control of Messenger RNA Stability, Academic Press, San Diego, CA, 1993, pp. 71–116.

[11] A.J. Carpousis, G. Van Houwe, C. Ehretsmann, H.M. Krisch, Copurification of *E. coli* RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation, Cell 76 (1994) 889–900.

[12] S. Naureckiene, B.E. Uhlin, In vitro analysis of mRNA processing by RNase E in the *pap* operon of *Escherichia coli*, Mol. Microbiol. 21 (1996) 55–68.

[13] A.W. Nicholson, Structure, reactivity and biology of double-stranded RNA, Prog. Nucleic Acid Res. Mol. Biol. 52 (1996) 1–65.

[14] G.A. Coburn, G.A. Mackie, Degradation of mRNA in *Escherichia coli*: an old problem with some new twists, Prog. Nucleic Acid Res. Mol. Biol. 62 (1999) 55–108.

[15] S.F. Newbury, N.H. Smith, E.C Robinson, I.D. Hiles, C.F. Higgins, Stabilization of translationally active mRNA by prokaryotic REP sequences, Cell 48 (1987) 297–310.

[16] S.F. Newbury, N.H. Smith, C.F. Higgins, Differential mRNA stability controls relative gene expression within a polycistronic operon, Cell 51 (1987) 1131–1143.

[17] R.S. McLaren, S.F. Newbury, G.S. Dance, H.C. Causton, C.F. Higgins, mRNA degradation by processive 3′–5′ exoribonucleases in vitro and the implications for prokaryotic mRNA decay in vivo, J. Mol. Biol. 221 (1991) 81–95.

[18] S. Dasgupta, L. Fernandez, L. Kameyama, T. Inada, Y. Nakamura, A. Pappas, D.L. Court, Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNase III—the effect of dsRNA binding on gene expression, Mol. Microbiol. 28 (1998) 629–640.

[19] J. Schultz, R.R. Copley, T. Doerks, C.P. Ponting, P. Bork, SMART: a web-based tool for the study of genetically mobile domains, Nucleic Acids Res. 28 (2000) 231–234.

[20] P. Regnier, M. Grunberg-Manago, RNase III cleavages in non-coding leaders of *Escherichia coli* transcripts control mRNA stability and genetic expression, Biochimie 72 (1990) 825–834.

ERIC repeats in *Yersiniae*: genomic organization and functional properties

running title ERIC elements in Yersiniae

Eliana De Gregorio, Giustina Silvestro, Mauro Petrillo,° Maria Stella Carlomagno and Pier Paolo Di Nocera*

Dipartimento di Biologia e Patologia Cellulare e Molecolare
Facoltà di Medicina, Università Federico II
Via S. Pansini 5, 80131 Napoli, Italy

° CEINGE Biotecnologie Avanzate S.C.a r. l.
Via Comunale Margherita n. 482, 80131 Napoli, Italy

*corresponding author
phone 0039-81-7462059
Email: dinocera@unina.it

Abstract

Genome-wide analyses carried out *in silico* revealed that the DNA repeats present in several enterobacteriaceae called ERICs (Enterobacterial Repetitive Intergenic Consensus sequences) are over-represented in *Yersiniae*. From the alignment of DNA regions from the wholly sequenced *Yersinia enterocolitica* 8081 and *Yersinia pestis* CO92 strains we could establish that ERICs are miniature mobile elements whose insertion leads to duplication of the dinucleotide TA. ERICs feature long terminal inverted repeats (TIRs) and can fold as RNA into hairpin structures. The proximity to coding regions suggests that most *Y.enterocolitica* ERICs are cotranscribed with flanking genes. Elements which either overlap, or are located next to stop codons, are preferentially inserted in the same (or B) orientation. In contrast, ERICs located far apart from ORFs are inserted in the opposite A orientation. The expression of genes cotranscribed with A- and B-oriented ERICs has been monitored *in vivo*. In mRNAs spanning B-oriented ERICs, upstream gene transcripts accumulated at lower levels than downstream gene transcripts. This difference was abolished by treating cells with chloramphenicol. We hypothesize that folding of B-oriented elements is impeded by translating ribosomes. Consequently, upstream RNA degradation is triggered by the unmasking of a site for the ribonuclease E located in the right-hand TIR of ERIC. A-oriented ERICs may act in contrast as upstream RNA stabilizers, or may have other functions. The hypothesis that ERICs act as regulatory RNA elements is supported by analyses carried out in Yersinia strains which either lack ERIC sequences, or carry alternatively oriented ERICs at specific *loci*.

Introduction

Transposable elements (TEs) are widely distributed in prokaryotic and eukaryotic genomes. According to their transposition intermediate, TEs are broadly divided into two classes. Class-1 elements transpose by means of an RNA intermediate and feature either long terminal direct repeats or a poly(A) tract at one end. Class-2 elements transpose by means of a DNA intermediate and most have terminal inverted repeats (TIRs). Integration of most TEs frequently determines the duplication of target sites of fixed lengths (20).

DNA repeats which recall class-2 elements for the presence of TIRs but have no coding capacity are found in many organisms. These non-autonomous mobile elements are commonly referred to as MITEs (Miniature Inverted Transposable Elements). First recognized as a predominant sequence type in plants, MITEs had been subsequently identified in many invertebrate and vertebrate genomes (14). A few MITE families have been characterized in archeal genomes (5, 34) and in eubacteria. *Streptococcus pneumoniae* contains ˜100 copies of a 107 bp long miniature insertion sequence called RUP (29). The 106-158 bp long DNA elements known as Correia or NEMIS make up 1-2% of the genome in pathogenic Neisseriae (6, 10, 22, 24). RUP and NEMIS feature similar TIRs, and both induce the duplication of the TA dinucleotide upon genomic insertion. Most NEMIS are cotranscribed with neighbouring genes, and NEMIS$^+$ mRNAs fold into hairpins formed by NEMIS *termini* which are targeted by RNAseIII (9, 11). Genome-wide analyses carried out *in silico* predict that the expression levels of 80-100 *N. meningitidis* genes may be tuned by RNAseIII-dependent processing at NEMIS RNA hairpins (10, 11).

The 127 bp long elements known either as IRU (38) or ERIC (17) structurally recall NEMIS and RUP repeats. ERIC families are made up by 20 to 30 elements in both *E. coli* and *S. typhimurium*. In this report, we show that ERICs, as anticipated by early genomic analyses by Hofnung and coworkers (3), are over-represented in *Yersiniae*. *In silico* analyses performed on the wholly sequenced *Yersinia pestis* CO92 (12, 30) and *Yersinia enterocolitica* 8081 (www.sanger.ac.uk/Projects/Y_*enterocolitica*) strains establish that ERIC constitute a major DNA family in Yersiniae. ERICs are (or have been) mobile DNA sequences which also belong to the MITE superfamily. Most of the 247 elements found in *Y. enterocolitica* are inserted at close distance from flanking coding regions, and it is likely that many are transcribed into mRNA. In this paper we show that, according to their orientation and relative position within the mRNA, transcribed ERICs may impede or accelerate the decay of specific mRNA segments.

Materials and Methods

Bacterial strains and growth conditions. The *Y. enterocolitica* strain Ye161 (serogroup 08) was kindly provided by Dr. Ida Luzzi at the Istituto Superiore di Sanità, Rome. The *Y. enterocolitica* strains Ye24 (serogroup 08) and Ye25 (serogroup 09), and the *Y. kristensii* SS47 strain were provided by Dr. Francesca Berlutti at the Istituto di Igiene of the University La Sapienza, Rome. Yersinia cells were grown in LB broth at 28°C. When needed, exponentially growing Ye161 cells were exposed either 12 min to rifampicin (200 micrograms/ml, final concentration) or 30 min to chloramphenicol (50 micrograms/ml, final concentration) before harvesting.

RNA analyses. Total bacterial RNA was purified on RNeasy column (Qiagen). Transcripts spanning the *cheW* (ORF YE2576), *trpB* (ORF YE2213), *uncE* (ORF YE4221 and *lpdA* (ORF YE0702) genes were monitored by RNA extension analyses as reported (9), by using as primers the pex.cheW, pex.trpB, pex.uncE and pex.lpdA oligonucleotides, respectively. The sequence of the four primers is reported in Table III. RT-PCR analyses were carried out by reverse transcribing 200 nanograms of total *Y. enterocolitica* RNA by random priming. The resulting cDNA was amplified by using pairs of gene-specific oligonucleotides (Table II). The Tm (melting temperature) of each oligomer (see Table II) was determined by using the Oligo 4,0 primer analysis software (35). In several instances RT-PCR co-amplifications were carried out with alternative pairs of primers. One oligonucleotide of each pair had been 32P-end-labelled at the 5' terminus with the polynucleotide kinase. Comparable yields of amplimers were obtained by labelling either forward or reverse cistron-specific primers. To adequately monitor gene-specific RNA levels by RT-PCR, the cDNA was amplified under non-saturating cycling conditions, and *ad hoc* low-cycles (6 to 12 cycles) PCR analyses were performed for each set of co-amplified genes. Amplimers were electrophoresed onto 6% polyacrylamide-8M urea gels, and quantitated by phosphorimagery.

For RNase protection assays, uniformly 32P-labelled RNA probes were obtained by transcribing *in vitro* linear DNA templates as described (9). Templates were obtained by PCR amplification of Ye161 DNA with the 45-mers shown in Table III. Within each pair, one oligomer included the sequence of the T7 RNA polymerase promoter in the 5' region. Twenty micrograms of total RNA were mixed with 32P-labeled anti-sense RNA probes in 30 microliters of hybridization buffer (75% formamide, 20 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl, 0.1% SDS). Samples were incubated at 95°C for 5 min, cooled down slowly and kept at 45°C for 16 h. After a 60 min incubation at 33°C with RNase T1 (2

micrograms/ml) samples were treated with proteinase K (50 micrograms/ml) for 15 min at 37°C, extracted once with phenol, precipitated with ethanol, resuspended in 80% formamide and loaded onto 6% poly-acrylamide/8M urea gels.

Computer analysis. *E. coli* ERIC sequences were used as queries in BLAST searches (2) to fetch homologous DNA segments from the genomes of the *Y. pestis* CO92 (30) and KIM (12) strains, and the *Y. enterocolitica* 8081 (www.sanger.ac.uk/Projects/Y_*enterocolitica*). Species-specific queries allowed the identification of Yersinia ERICs evolutionarily distant from *E. coli* homologs. Retrieved DNA sequences were aligned with the CLUSTAL W program (41). Consensus sequences from multiple alignments of ERIC family members were established with the program CONS of the EMBOSS package. Secondary structure modeling was done using the Mulfold program (www.bioinfo.rpi.edu/applications/mfold) which predicts RNA secondary structure by free energy minimization (45).

Results

Genomic and structural organization of ERICs in Yersiniae

The genus *Yersinia* includes eleven species, three of which are pathogenic to man (4). The enteropathogens *Y. pseudotuberculosis* and *Y. enterocolitica* are widely found in the environment. In contrast, *Y. pestis* is a highly virulent blood borne pathogen transmitted by fleas, which rapidly evolved from *Y. pseudotuberculosis* (1, 44). *In silico* analyses carried out on wholly sequenced strains showed that Yersinia chromosomes are peppered by ERIC repeats. By looking only at elements carrying both TIRs (see Fig. 1) we found 247 and 167 ERICs in the genome of the *Y. enterocolitica* 8081 and *Y. pestis* CO92 strains, respectively (Table I). About 90% of the elements are scattered throughout the chromosome of either species as single copy insertions. The remaining 10% is made up by clusters in which two to five elements are organized in head-to-tail configuration. On the whole, 235 and 151 ERIC+ sites were identified in Y. *enterocolitica* 8081 and Y. *pestis* CO92 strains, respectively (Table I). In contrast, the *Y. pestis* CO92 strain contains several moderately abundant families of IS (65 copies of IS1541, 44 copies of IS100, 8 copies of IS1661 and 21 copies of IS285; see ref 7), while we found only 3 copies of IS1541 in the Y. *enterocolitica* 8081 genome by BLAST analyses (not shown).

Unit-size ERICs are 127 bp in length (Fig. 1). Shorter elements measure ˜70 bp, and all lack a 50 bp long internal segment. Larger elements are interrupted at specific sites by three different types of DNA insertions (Fig. 1). Type-1 and type-2 insertions had been found also in some E. *coli* ERICs (37), type-3 insertions seem to be present only in Yersiniae.

Y. pestis and *Y. enterocolitica* genomes both measure ˜4,6 Mb. Yet, extensive genetic remodeling makes *Y. pestis* a species evolutionarily distant from other Yersiniae (44). *Y. pestis* ERICs are fewer and exhibit more size heterogeneity than *Y. enterocolitica* elements (Fig. 1A). The *Y. pestis* CO92 and the *Y. enterocolitica* 8081 chromosomes share only 37 syntenic regions carrying ERIC repeats. Elements have the same size only in 1/3 of the cases. In the other instances, unit-length elements found in *Y. pestis* are replaced by either shorter, or insertions-tagged ERICs in *Y. enterocolitica*, and *viceversa* (not shown), plausibly as a result of recombination events between ERIC family members.

The insertion of ERIC induces a 2-bp target site duplication

Several syntenic regions identified in *Y. enterocolitica* and *Y. pestis* carry an ERIC element in the former species only. ERICs terminate at either side with the dinucleotide TA. At many *Y. pestis* empty sites, ERIC is replaced by one copy of the dinucleotide (Fig. 2). The duplication of the dinucleotide TA is an hallmark of eukaryotic MITEs, and is a feature shared by known eubacterial MITEs (24, 29). TA

empty sites have been identified both in *Y. enterocolitica* and *Y. pestis* (Table I). This indicates that the mobilization of ERICs still occurred after the speciation of Yersiniae into *Y. enterocolitica* and *Y. pseudotuberculosis*, from which *Y. pestis* eventually derived (1, 44).

Differences in the distribution of ERICs between the *Y. pestis* CO92 and *Y. enterocolitica* 8081 genomes reflect species-specific, rather than strain-specific variations. The *Y. pestis* strains CO92 (30) and KIM (12) belong to different biovars, have been responsible for the spread of different plague epidemics and show a remarkable amount of genome rearrangement (12). However, 155/157 ERIC+ sites found in the CO92 strain are perfectly conserved, as revealed by BLAST analyses, in the KIM strain, and the remaining 2 sites vary only for the number of tandemly inserted elements.

ERICs are co-transcribed with flanking genes

Genome-wide surveys revealed that 137 ERICs are inserted at close distance (≤50 bp) from either the start or the stop codons of *Y. enterocolitica* 8081 ORFs (not shown). This suggests that most elements are co-transcribed with flanking genes into mRNAs.

To investigate on the issue, we first checked that ERIC+ regions found in the 8081 strain were conserved in the *Y. enterocolitica* Ye161 strain. Subsequently, the corresponding ERIC+ mRNAs synthesized in this strain were monitored by primer extension analyses (Fig. 3). The major products of extension of both *lpdA* and *uncE* transcripts extended beyond ERIC (Fig. 3). In contrast, extension products of both *cheW* and *trpB* transcripts were found to terminate at multiple sites within ERIC sequences (Fig. 3). The same pattern was obtained with different RNA preparations and reverse transcriptase batches. The multiple extension products detected may denote cleavage of *cheW* and *trpB* mRNAs at ERIC sequences.

ERICs flanking the *lpdA* and *uncE* genes are both inserted in the same orientation, that herein on we will arbitrarily refer to as A orientation. In contrast, elements flanking the *cheW* and *trpB* genes are inserted in the alternative B orientation. About 110 ERICs are found, within a -50/+45 bp distance range, downstream from the stop codons of annotated ORFs in the 8081 strain (Fig. 4). Curiously, most ERICs which either overlap, or are inserted next (0/+6 bp distance range) to stop codons, are B-oriented elements. In contrast, ERICs located at larger distances from ORFs are predominantly A-oriented elements (Fig. 4). The orientation-dependence rule works for elements located between unidirectionally transcribed ORFs as for elements separating convergently transcribed ORFs. A priviliged orientation relatively to the distance from translational stop codons was similarly displayed by ERICs found in the *Y. pestis* CO92 strain (not shown).

To investigate on the functional significance of these observations, several pairs of *Y. enterocolitica* genes transcribed in the same direction, but separated by either A- or B-oriented ERICs, were selected for comparative RNA quantitations. Elements analyzed measured all 127 bp and exhibited 94% sequence similarity. Total RNA from Ye161 cells was reverse transcribed into cDNA, an the latter was subsequently amplified by using different sets of primers. As evidenced by the detection of large mRNA segments, elements selected are cotranscribed with flanking genes (Fig. 5A). To monitor the relative abundance of RNA species corresponding to upstream and downstream cistrons, the cDNA was co-amplified with pairs of cistron-specific oligomers (Table II) under non-saturating cycling conditions (Fig. 5B). Radiolabeled amplimers were separated electrophoretically, and their amount quantitated by phosphorimagery. Data obtained with alternative sets of primers were fairly comparable, ruling out

technical artifacts. By looking at trascriptional units spanning B-oriented ERICs, we found that downstream gene transcripts accumulated ˜4-fold more abundantly than upstream gene transcripts. In contrast, except for the *glgC*/*glgA* barrier, the levels of gene transcripts flanking A-oriented ERICs were comparable (Fig. 5C). Differences in the downstream/upstream gene transcript ratios measured at intercistronic barriers carrying A-oriented (*panB -panC*) and B-oriented (*cheY -cheB*) ERICs were confirmed by RNase protection experiments, and magnified when *de novo* RNA synthesis was blocked by treating Yersinia cells with rifampicin (Fig. 6). Both *panB* and *panC* transcripts, which are quite abundant in steady-state RNAs, were no longer detected after exposure of *Y. enterocolitica* cells to rifampicin (compare lanes 20 and 21 in Fig. 6B). By contrast, the difference in the steady-state levels of *cheY* and *cheB* transcripts made still possible to detect *cheY* RNA sequences in rifampicin-treated cells (compare lanes 9 and 10 in Fig. 6A).

Data signal that the segmental stability of RNAs spanning A-oriented and B-oriented elements was substantially different.

Heterogenity of ERIC+ loci among Yersiniae

The conservation of ERIC sequences in *Y. enterocolitica* was monitored by PCR-driven surveys. The Ye161 and the Ye24 strains, and the sequenced 8081 strain, all belong to the 08 serogroup. It is therefore not surprising that 24/24 ERIC+ loci analyzed (including those in Fig. 5) were conserved in the three strains (data not shown). In contrast, genetic variations at specific *loci* spanning ERIC sequences found in the 8081 strain were identified in Ye25, a 09 serogroup *Y. enterocolitica* strain, as in the YkSS47 strain of the apathogenic *Yersinia kristensii* species, and exploited for comparative RNA analyses. In Ye161, *cheA* and *cheW* genes are separated by a B-oriented ERIC, and *cheW* transcripts are ~5-fold more abundant than *cheA* transcripts. The difference is abolished in YkSS47 cells (Fig. 7). Sequence analysis showed that the YkSS47 *cheA-cheW* region did not experienced the insertion of ERIC DNA. In Ye161, *argB* and *argH* genes are separated by a B-oriented ERIC inserted immediately dowstream from the *argB* stop codon. In contrast, in Ye25 the two genes are separated by an A-oriented ERIC inserted 10 bp downstream from the *argB* stop codon. Changes in the position and the orientation of ERIC are associated to significant differences in the *argH-argB* transcripts ratios (Fig. 7). Finally the ERIC which separates *panB* and *panC* genes in Ye161 is missing in the YkSS47 strain. This correlates with a 3-fold decrease in the level of the *panB* transcripts (Fig. 7).

Translating ribosomes and RNA folding

Data shown in Figs. 5-7 support the notion that the relative abundance of the mRNA segments flanking ERIC sequences may be influenced by the orientation of ERICs. The high downstream/upstream transcript ratio measured at intercistronic barriers spanning B-oriented elements may correlate with the activity of promoter sequences directing the synthesis of transcripts toward downstream genes. In A-oriented ERICs the hypothetical promoter would also direct the synthesis of transcripts toward upstream genes, causing transcriptional collisions and allowing for the formation of antisense RNA. It is difficult to envisage how this may be advantageous to the organism. Moreover, it is left unexplained why B-oriented elements tend to be inserted so close to stop codons. We rather believe that transcribed ERICs may act as modulators of RNA decay, and that A- and B-oriented elements may function in different ways. According to this hypothesis, the high downstream/upstream gene transcript ratios measured at intercistronic barriers carrying B-oriented ERICs may be the result of processing events promoted by ERIC repeats which enhance upstream RNA degradation.

The orientation-dependent mode of action suggests that a sequence must be crucial for upstream RNA instability. RNAs corresponding to A-oriented and B-oriented ERICs may fold into secondary

structures which have similar shapes and comparable calculated free energies (Fig. 8A; see refs. 17 and 38). The formation of RNA hairpins is preserved in the majority of elements by compensatory mutations, and is unaffected in shorter as well as larger ERICs, because both type-1 and type-2 insertions feature self-complementary regions (Fig. 1D; see also ref. 37). However, the left-hand TIRs of ERICs which are inserted close to stop codons are covered by terminating ribosomes, a translating ribosome protecting at least 30 residues of the mRNA (40). Noteworthy an AU-rich sequence (AAUUAUUUA, see Fig. 8A) would not be base-paired in B-oriented elements because of steric hyndrance caused by ribosomes. Unfolded AU-rich sequences represent preferred cleavage sites for RNase E (13, 19, 21, 26). The enzyme, which is conserved both in *Y. enterocolitica* and *Y. pestis* (ORFs YE1627 and YPO1590, respectively), is the major endoribonuclease responsible for the mRNA decay in bacteria (8), and is associated in E. *coli* to the 3'-5' exoribonucleases polynucleotide phosphorylase (PNPase) and RNAseII, in the molecular machine known as degradosome (8, 32). The mRNA degradation by 3'-5' exonucleases subsequent to RNase E-mediated cleavage may explain the high downstream/upstream transcript ratios measured at specific ERIC$^+$ intercistronic barriers (Figs. 6-8).

Experimental support to this hypothesis is provided by data shown in Fig. 8B. Cleavage of ERIC$^+$ mRNAs should be favoured by the occupancy of the left-hand ERIC TIR by translating ribosomes. Moreover, uncoupling transcription and translation should alter the dowstream/upstream gene transcripts ratio in ERIC$^+$ mRNAs spanning B-oriented ERICs only. ERICs located downstream from the *cheA* and *panB* genes are inserted in the B- and A-orientation, respectively (Fig. 5). Treatment of Ye161 cells with chloramphenicol significantly altered the *cheW*/*cheA* transcripts ratio, as we measured a four-fold increase in the amount of *cheW* RNA, but had no effect on the *panC*/*panB* transcripts ratio (Fig 8B; see refs. 23 and 39).

Noteworthy, the predominant extension products corresponding to the **e** and **l** bands in Fig. 3 nicely match in size RNA species respectively generated by cleavage of *cheW* and *trpB* transcripts at the AU-rich site within the upstream B-oriented ERICs.

A-oriented ERICs are found far from stop codons, and therefore can fold into RNA hairpins. These elements may therefore act in an opposite way to B-oriented ERICs, functioning as upstream RNA stabilizers (see Discussion).

Discussion

Origin and evolution of ERIC sequences

ERIC repeats are present in several bacterial species as low-copy number families, and PCR fingerprinting using ERIC primers is widely used for diagnostic purposes (43). In contrast, ERICs are a major genome component in pathogenic *Yersiniae*, accounting for ˜0.7% and ˜0.45% of the *Y. enterocolitica* and *Y. pestis* total DNA content, respectively. In either species, elements are scattered throughout the chromosome mostly as single copy insertions. The genomic spread of ERICs occurred most probably by transposition. As unambigously set by the comparison of empty and filled chromosomal sites, ERICs specifically duplicate the dinucleotide TA upon genomic insertion (Fig. 2). This is an hallmark of miniature transposable elements originating from members of the IS630-Tc1-mariner TE superfamily known as MITEs. The mobilization of ERICs, initially fostered by large codogenic progenitor ISs, might have been eventually mediated, as suggested for eukaryotic MITEs (15, 18, 31, 36), also by ISs whose transposases were able to recognize. ERIC are plausibly no longer mobile in Yersiniae, as we couldn't identify *in silico* neither *bona fide* ERIC progenitor nor potential cross-mobilizing TEs neither in the sequenced *Y. enterocolitica* and *Y. pestis* strains, nor in the genome of the *Y. pseudotuberculosis* IP32593 strain, whose sequence has been recently determined (7). Data reported in this work support the notion that Yersiniae learned during evolution to exploit the genomic spread of ERICs for functional purposes.

ERICs as modulators of RNA decay

In Yersiniae, ERICs are frequently inserted next to codogenic regions, and most are plausibly transcribed into mRNAs. The ability of ERIC RNA to fold into relatively robust, low free energy RNA hairpins (Fig. 8A), is a feature previously noted (17, 38).

Whole *in silico* surveys surprisingly revealed a privileged orientation of ERIC sequences relatively to their position in the mRNA. In the *Y. enterocolitica* 8081 strain, 56/60 elements which either overlap, or are located 6 bp or less far from the stop codon of annotated ORFs, are inserted in the same orientation (B-oriented ERICs). By contrast, 45/47 elements located more distantly from stop codons (distance range, +7/+35) are inserted in the opposite orientation (A-oriented ERICs). This peculiar organization must convey a selective advantage in evolution for functional purposes.

The preferential location next to stop codons implies that RNA hairpins formed by B-oriented ERICs are remodelled by terminating ribosomes (Fig. 8C). We hypothesize that inhibiting secondary structure

formation unmasks a potential target site for RNase E which is located in the right-end TIR of these elements. In turn, the endonucleolytic cleavage activates the degradation of upstream RNA segments by the polynucleotide phosphorylase (PNPase) and RNAseII, the two 3'-5' exoribonucleases associated to the RNase E in the degradosome (8, 32).

Translation should not interfere with the formation of RNA secondary structures in A-oriented ERICs. By folding into stable RNA hairpins these repeats should be able to slow down the degradation of upstreamly located RNA segments by impeding the passage of 3'-5' exonucleases. These repeats may thus work analogously to the shorter intergenic sequences found in E. coli know as REPs (16). The element found at the *glgC/glgA* intercistronic barrier seems to work this way (Fig. 5). Similar conclusion can be reached for the element found between *panB* and *panC* cistrons (Fig. 7). Yet, in other transcriptional units spanning A-oriented elements, upstream and downstream trascripts accumulated at similar levels (Fig. 5). We do not have an explanation for such discrepancies. Plausibly several A-oriented ERICs cannot function as upstream RNA stabilizers, because they are overriden by dominant instability determinants located in the mRNA. Such a phenomenon has been documented for different E. *coli* REPs (25, 27, 28). Similarly, the degradation of 5' flanking RNA prompted by B-oriented ERICs may be impaired by mRNA stability determinants. The efficacy by which ERICs modulate RNA decay may be vary not only because of the intrinsic stabilities of neighbouring mRNA segments, but also because of sequence heterogeneity among ERICs. Thus, conclusions on the ability of members of the ERIC family to function as RNA control elements can be drawn in many instances only by integrating sequence data with functional RNA analyses.

In spite of the smaller size of the family, also *Y. pestis* ERICs can be largely sorted into A-oriented and B-oriented elements according to their distance from upstream ORFs. Whether the ERIC-dependent modulation of RNA decay works in this species, which rapidly evolved as an arthropode-adapted pathogen, remains to be established.

In the *Y. enterocolitica* 8081 strain, 30 elements are inserted relatively far from ORF stop codons, but close (≤50 bp distance) to ORF start codons. These repeats may either stabilize downstream RNA sequences (see *lpdA* and *uncE* transcripts in Fig. 3), or interfere with mRNA translation. Some ERICs could alternatively function as DNA, rather than as RNA elements. Yet, deleting an ERIC from the promoter region of the *Y. enterocolitica cpdB* gene had no effect on *cpdB* expression (42). By contrast, the ERIC found in the promoter of the *Y. enterocolitica ybtA* yersiniabactin regulator may modulate yersiniabactin activity, as putative binding sites for the YbtA transcriptional regulator and the TATACCC motif found in ERIC TIRs coincide (33).

The number, the structural organization and the chromosomal distribution of ERICs and neisserial NEMIS sequences are similar. It is curious noting that members of these two MITE families, spread in evolutionarily distant Gram$^-$ bacteria, independently evolved into substrates for the major cellular endoribonucleases. We wouldn't be surprised to learn that bacterial MITEs yet to be discovered may have similarly evolved into cis-acting sequences regulating mRNA metabolism. Whether MITE-like repeats found in eukaryotes may similarly work as RNA regulatory elements remains to be established.

Acknowledgements

We are indebted to Drs. Ida Luzzi and Francesca Berlutti for providing us with Yersinia strains, and to Prof. Bruno Bruni for critical revision of the manuscript. This work has been funded by a grant assigned to Pier Paolo Di Nocera by the PRIN 2004 agency of the Italian Ministry of the University and Scientific Research.

References

1. **Achtman, M., K. Zurth, , G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel.** 1999. Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. USA. **96:** 14043-14048.

2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:** 403-410.

3. **Bachellier, S., J. M. Clement, and M. Hofnung.** 1999. Short palindromic repetitive DNA elements in enterobacteria, a survey. Res. Microbiol. **150:** 627–639.

4. **Bottone, E. J.** 1999. Yersinia enterocolitica, overview and epidemiologic correlates. Microbes Infect. **1:** 323-333.

5. **Brugger, K., P. Redder, Q. She, F. Confalonieri, Y. Zivanovic, and R. A. Garrett.** 2002. Mobile elements in archaeal genomes. FEMS Microbiol Lett. **206:** 131-141.

6. **Buisine, N., C. M. Tang, and R. Chalmers.** 2002. Transposon-like Correia elements, structure, distribution and genetic exchange between pathogenic Neisseria species. FEBS Lett. **522:** 52-58.

7. **Chain, P. S., E . Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser, E. Garcia. 2004.** Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. U S A. **101:**13826-13831.

8. **Coburn, G. A., and G. A. Mackie.** 1999. Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. Prog. Nucleic Acid Res. Mol. Biol. **62:** 55-108.

9. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2002. The abundant class of nemis repeats provides RNA substrates for ribonuclease III in Neisseriae. Biochim Biophys Acta **1576:** 39-44.

10. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2003. Asymmetrical distribution of Neisseria miniature insertion sequence DNA repeats among pathogenic and nonpathogenic Neisseria strains. Infect. Immun. **71:** 4217-4221.

11**. De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2003. Ribonuclease III-mediated processing of specific Neisseria meningitidis mRNAs. Biochem J. **374:** 799-805.

12. **Deng, W., V. Burland, G. 3rd. Plunkett, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles. J. S. Matson, F. R. Blattner, and R. D. Perry.** 2002. Genome sequence of Yersinia pestis KIM. J Bacteriol. **184:** 4601-4611.

13. **Ehretsmann, C. P., A. J. Carpousis, and H. M. Krisch.** 1992. Specificity of Escherichia coli endoribonuclease RNase E, *in vivo* and *in vitro* analysis of mutants in a bacteriophage T4 mRNA processing site. Genes Dev. **6:** 149-159.

14. **Feschotte, C., N. Jiang, and S. R. Wessler.** 2002. Plant transposable elements: where genetics meets genomics. Nat Rev Genet. **3:** 329-341.

15. **Feschotte, C., L. Swamy, and S. R. Wessler.** 2003. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). Genetics **163:** 747-758.

16. **Higgins, C. F., R. S. McLaren, and S. F. Newbury.** 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression, evolution by gene conversion? A review. Gene **72:** 3-14.

17. **Hulton C. S. J., C. F. Higgins, and P. M. Sharp.** 1991. ERIC sequences, a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. **5:** 825-834.

18. **Jiang, N., C. Feschotte, X. Zhang, and S. R. Wessler.** 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol. **7:** 115-119.

19. **Kaberdin, V. R.** 2003. Probing the substrate specificity of Escherichia coli RNase E using a novel oligonucleotide-based assay. Nucleic Acids Res. **31:** 4710-4716.

20. **Kidwell, M. G., and D. R. Lisch.** 2000. Transposable elements and host genome evolution. Trends Ecol. Evol. **15:** 95-99.

21. **Lin-Chao, S., T. T. Wong, K. J. McDowall, and S. N. Cohen.** 1994. Effects of nucleotide sequence on the specificity of rne-dependent and RNase E-mediated cleavages of RNA I encoded by the pBR322 plasmid. J Biol Chem. **269:** 10797-10803.

22. **Liu, S. V., N. J. Saunders, A. Jeffries, and R. F. Rest.** 2002. Genome analysis and strain comparison of correia repeats and correia repeat-enclosed elements in pathogenic Neisseria. J. Bacteriol. **184:** 6163-6173.

23. **Lopez, P. J., I. Marchand, O. Yarchuk, and M. Dreyfus.** 1998. Translation inhibitors stabilize Escherichia coli mRNAs independently of ribosome protection. Proc. Natl. Acad. Sci. U S A. **95:** 6067-6072.

24. **Mazzone, M., E. De Gregorio, A. Lavitola, C. Pagliarulo, P. Alifano, and P. P. Di Nocera.** 2001. Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic Neisseriae. Gene **278:** 211-222.

25. **McCarthy J. E., B. Gerstel, B. Surin, U. Wiedemann, P. Ziemke.** 1991. Differential gene expression from the Escherichia coli atp operon mediated by segmental differences in mRNA stability. Mol Microbiol. **10:** 2447-2458.

26. **McDowall, K. J., S. Lin-Chao, and S. N. Cohen.** 1994. A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. J Biol Chem. **269:** 10790-10796.

27. **Meyer B. J., and J. L. Schottel.** 1992. Characterization of cat messenger RNA decay suggests that turnover occurs by endonucleolytic cleavage in a 3' to 5' direction. Mol Microbiol. **9:** 1095-1104.

28. **Meyer B. J., A. E.. Bartman, and J. L. Schottel.** 1996. Isolation of a mRNA instability sequence that is cis-dominant to the ompA stability determinant in Escherichia coli. Gene. **179:** 263-270.

29. **Oggioni, M. R., and J. P. Claverys.** 1999. Repeated extragenic sequences in prokaryotic genomes, a proposal for the origin and dynamics of the RUP element in Streptococcus pneumoniae. Microbiology **145:** 2647-2653.

30. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Genome sequence of Yersinia pestis, the causative agent of plague. Nature **413:** 523-527.

31. **Plasterk R. H., Z. Izsvak and Z. Ivics.** 1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet. **15:** 326-332.

32. **Py, B., C. F. Higgins, H. M. Krisch, and A. J. Carpousis.** 1996. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. Nature **38:** 169-172.

33. **Rakin, A., C. Noelting, S. Schubert, and J. Heesemann.** 1999. Common and specific characteristics of the high-pathogenicity island of Yersinia *enterocolitica*. Infect. Immun. **67:** 5265-5274.

34. **Redder, P., Q. She, and R. A. Garrett.** 2001. Non-autonomous mobile elements in the crenarchaeon Sulfolobus solfataricus. J. Mol. Biol. **306:** 1-6.

35. **Rychlik, W., and R. E. Rhoads.** 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res. **17:** 8543-8551.

36. **Shao, H., and Z. Tu.** 2001. Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics **159:** 1103-1115.

37. **Sharp, P. M.** 1997. Insertions within ERIC sequences. Mol. Microbiol. **24:** 1314-1315.

38. **Sharples, G. J., and R. G. Lloyd.** 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosome. Nucl. Acids Res. **18:** 6503-6508.

39. **Sousa, S., I. Marchand, M. Dreyfus.** 2001. Autoregulation allows Escherichia coli RNase E to adjust continuously its synthesis to that of its substrates. Mol. Microbiol. **42:** 867-878.

40. **Steitz, J. A.** 1969. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature **224:** 957-964.

41. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673-4680.

42. **Trulzsch, K., A. Roggenkamp, C. Pelludat, A. Rakin, C. Jacobi, and J. Heesemann.** 2001. Cloning and characterization of the gene encoding periplasmic 2',3'-cyclic phosphodiesterase of Yersinia *enterocolitica* O 8. Microbiology **147:** 203-213.

43. **Versalovic, J., T. Koeuth, and J. R. Lupski.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. **19:** 6823-6831.

44. **Wren, B. W.** 2003. The Yersiniae, a model genus to study the rapid evolution of bacterial pathogens. Nat Rev Microbiol. **1:** 55-64.

45. **Zuker, M.** 1989. On finding all suboptimal foldings of an RNA molecule. Science **244:** 48-52.

Legends to the Figures

FIGURE 1 Eric elements in Yersiniae **A**) ERIC size classes in *Y. enterocolitica* 8081 and *Y. pestis* CO92 strains The number of elements carrying both TIRs found and their size in base pairs is indicated **B)** Structural organization of ERICs. Boxed arrows mark the terminal inverted repeats or TIRs. Triangles mark type-1 to type-3 insertions interrupting ERIC sequences **C)** The consensus sequence of the 127 bp unit-length ERICs is shown in the A-orientation. TIR residues are in capital letters. Underlined residues mark sequences conserved in the internally rearranged 70 bp long ERICs. The integration sites of type-1 to type-3 insertions are denoted by asterisks **D)** The consensus sequences of the three types of intervening sequences found to interrupt ERICs are shown.

FIGURE 2 Filled and empty ERIC sites. Homologous DNA regions from the *Y. enterocolitica* 8081 (Ye) and *Y. pestis* CO92 (Yp) strains are aligned. Numbers refer to genome residues, dashes denote sequence identities. The duplication of the TA target site at ERICs *termini* is highlighted.

FIGURE 3 Primer extension analyses of ERIC+ transcripts. 32P-5´-end-labelled primers complementary to *lpdA*, *uncE*, *cheW* and *trpB* transcripts were hybridized to total RNA (5 micrograms) derived from the *Y. enterocolitica* Ye161 strain. Annealed primer moieties were extended, in the presence of NTPs, by AMV (avian myeloblastosis virus) reverse transcriptase. Reaction products were electrophoresed on 6% polyacrylamide/8 M urea gels. Major reaction products labeled **a** to **l** are marked by arrows. Numbers to the left of each autoradiogram refer to the size in nucleotides of co-electrophoresed DNA molecular size markers. In the diagrams at the bottom is sketched the organization of the ERIC+ regions analyzed. The direction of transcription of the genes analyzed is indicated by dotted arrows. Primers are denoted by arrows, lines labeled **a** to **l** denote the extended products. Numbers refer to the distance in bp separating ERICs from either the stop or the start codons of neighbouring ORFs.

FIGURE 4 Asymmetry in the orientation of ERICs. The distances in base pairs separating B-oriented and A-oriented ERICs from flanking upstream ORFs in the *Y. enterocolitica* 8081 strain are indicated.

FIGURE 5 RT-PCR analyses of ERIC+ transcripts. Total RNA (200 nanograms) derived from the Ye161 strain had been reverse transcribed by using a mixture of random examers as primers. The cDNA obtained had been amplified by PCR with cistron-specific oligomers. One oligonucleotide of each pair of primers was 32-P-end-labelled to allow amplimer detection by autoradiography. Reaction products were run on a 6% polyacrylamide-8M urea gels. **A)** Transcripts spanning ERIC sequences and *cheA-cheW* (lane 1), *cheB-cheY* (lane 2), *manX-manY* (lane 3), *panB-panC* (lane 4), *trpC-trpB* (lane 5) and *pstS-pstC* (lane 6) genes were detected by using the external primers 1 and 4 under standard PCR cycling conditions (20-22 cycles). Amplimers were detected only when RNA samples were incubated with reverse transcriptase (+ lanes) prior to PCR. **B)** Total cDNA from the Ye161 strain had been amplified by using the 1, 2 and 3, 4 pairs of ORF-specific primers shown at the top for a limited number (6 to 12) of PCR cycles. Amplimers were quantitated by phosphorimaging. In the example reported, amplimers 1 and 2 correspond to the *cheA* and *cheW* genes, respectively **C)** The listed genes flanking ERIC repeats have been analyzed as described above. Distances in bp separating ERICs *termini* from stop and start codons of flanking ORFs are indicated. The orientation of ERIC (A, B) is given in parenthesis. The amount of transcripts corresponding to downstream (dw) and upstream (up) genes for each pair is expressed as ratio. RT-PCR analyses were routinely repeated three to four times on two independent RNA preparations. Standard deviations are indicated. The YE number assigned by the Sanger Centre to the ORFs analyzed, and their hypothesized function is listed below: cheA (YE2577), chemotaxis protein CheA; *cheW* (YE2576), chemotaxis protein CheW; *trpC* (YE2212), tryptophan biosynthesis bifunctional protein; *trpB* (YE2213), tryptophan synthase subunit B; *phoT* (YE4198), high-affinity P-specific transport, cytoplasmic ATP-binding protein; *phoU* (YE4196) P uptake, high-affinity P-specific transport system, regulatory gene; *cheB* (YE2571), glutamate methylesterase; *cheY* (YE2570), chemotaxis protein CheY; *glgC* (YE4011), glucose-1-phosphate adenylyltransferase; *glgA* (YE4010), glycogen synthase; *manX* (YE1777) mannose phosphotransferase system, EIIAB component; *manY* (YE1776) mannose phosphotransferase system, EIIC component; *panB* (YE0720) ketopantoate hydroxymethyltransferase; *panC* (YE0719), pantothenate synthetase; *pstS* (YE4201), high-affinity P-specific transport, periplasmic; *pstC* (YE4200) high-affinity P-specific transport; cytoplasmic membrane component.

FIGURE 6 RNase protection of ERIC+ transcripts. Uniformly 32P-labeled antisense RNA probes, complementary to the coding regions of the *Y. enterocolitica cheB-cheY* and *panB-panC* genes were transcribed *in vitro* by the T7 RNA polymerase. In the diagram RNA probes are sketched (not in scale)

below genes. Thicker segments mark complementarity to mRNA. Probes were hybridized to 20 micrograms of total RNA from Ye 161 cells untreated or exposed to rifampicin (200 micrograms/ml final concentration) for 12 min before cell harvesting. T1 RNase-resistant RNA hybrids were electrophoresed on 6% polyacrylamide/8M urea gels. Reaction products corresponding to *cheY*, *cheB*, *panC* and *panB* RNAs are marked by arrows. **A**) Analysis of *cheB-cheY* RNAs. Lanes: unreacted input probes (1, *cheY*; 5, *cheB*), probes hybridized separately to total RNA from Ye 161 cells untreated (2, *cheY*; 6, *cheB*) or exposed for 12 min to rifampicin (3, *cheY*; 7, *cheB*) or *E. coli* tRNA (4, *cheY*; 8, *cheB*). Probes were hybridized together to total RNA from Ye 161 cells untreated (9) or exposed for 12 min to rifampicin (10), or to *E. coli* tRNA (11). **B)** Analysis of *panB-panC* RNAs. Lanes: unreacted input probes (12, *panC*; 16, *panB*), probes hybridized separately to total RNA from Ye 161 cells untreated (14, *panC*; 18, *panB*) or exposed for 12 min to rifampicin (15, *panC*; 19, *panB*), or E. coli tRNA (13, *panC*; 17, *panB*). Probes were hybridized together to total RNA from Ye 161 cells untreated (20) or exposed for 12 min to rifampicin (21) or to *E. coli* tRNA (22).

Figure 7 Comparison of loci carrying or missing ERIC sequences in different Yersinia strains. ERIC elements are depicted as grey boxes, and number within refer to element sizes. Number above boxes signal the distance in bp separating ERIC from flanking ORFs. Total RNA derived from the *Y. enterocolitica* Ye161 and Ye25, and the *Y. kristensii* YkSS47 strains was analyzed by RT-PCR as in Fig. 5. At the empty genomic sites identified in the genome of the YkSS47 strain, ERIC sequences are replaced by the dinucleotide TA.

FIGURE 8 A) Predicted RNA foldings and relative calculated free energies at 37°C of unit-length ERIC consensus sequences inserted in A- and B- orientation. Non Watson-Crick base pairings are highlighted by dots. The hypothesized cleavage site for RNase E, present in B-oriented ERICs, is boxed **B)** Translation-dependent processing of ERIC+ RNAs. Total RNA derived from exponentially growing Ye161 cells untreated (-) or exposed (+) for 30 min to chloramphenicol (50 microgams/ml final concentration) was analyzed as in Fig. 5. **C)** Ribosomes interfere with folding of ERIC+ RNA. In mRNA spanning B-oriented elements that are inserted close to the translational stop codon, the translating ribosome covers most of the ERIC left-hand TIR, unmasking the RNaseE site (sketched as a triangle) located in the ERIC right-hand TIR.

	Ye	Yp
copy number	247	167
ERIC+ sites	235	151
single-copy inserts	225	148
type-I insertions	13	11
type-II insertions	12	2
type-III insertions	5	35
TA empty sites	23	3

Table I

ERIC elements in wholly sequenced *Y. enterocolitica* 8081 (Ye) and *Y. pestis* CO92 (Yp) strains.

gene	direct primer	Tm(°C)	reverse primer	Tm(°C)
cheW	CGAAACGGTAGGACAAGAATTCCTG	59,1	GGAACAATAACGCCGCGTAAGTTAG	59,2
cheA	TCATTTTACCATTGAACGCCGTAAT	57,7	CAGAAATACCTGGAACCTTGCGATA	57,5
cheA	AGGCATTGTTGTGATTCTACAAAGC	55,7	ACGACTCCATCATCAGCCGTAACAC	60,2
trpC	TTGAACGCTATGTTTTGGATAATGG	56,3	CAATCTTTTGGGGATCTTTAATGCC	58,1
trpB	ATCCCTATTTTGGCGAGTTCGGGGG	56,2	CAGAGCGGTTGGACGCCCAGCATAG	58,0
phoU	CATATTTCCGGCCAGTTCAATGCAG	62,3	CTTTTTATCACCCTCGATGACGCGC	63,5
phoT	GTTATAGCTTGTCGGGTGGGCAGCA	63,9	TTCTGCTGTGGTGCGGTAAACAGGG	64,8
phoT	AGGTATCGCCATTCGTCCAGATGTG	62,0		
cheY	TGGTAGACGATTTTTCGACCATGCG	62,8	TCGGCATGTTCCAGTCAGAAACCAC	63,0
cheB	AACTATCAGGTGCGTATTCATGATG	64,2	GCTTCGTTTTGTGCAATGGTATAAG	66,2
cheB	AATTACCGTGAAGGAAGCAGAAGAC	66,2		
glgA	CGTATGTTCTGAGCTATTCCCGTTG	58,1	AGCAAAGGTATCAATCTCCCTGACC	57,9
glgA	TACGGCATGGAGGGGTTGCTACAAG	60,5	AAGCCATACAACTGTGTTAGCCCAC	57,5
glgC	GTAGCCACGGTATGACCATGAACTC	57,7	GCAGCAGTAATGTGGAATCAATGGT	58,2
glgC			GATAAAAACGCTTGCTGTCTTCTTC	55,0
manX	GCAATATGCCAAAGACCGGGTCATG	64,0	TTCGGACTTCCAGTTCAATACCGCG	63,9
manY	ATTTCATCGTTGCCTGTATCGCCGG	63,7	ACCAAAGTACAGGCGACGAGGGGAC	64,1
manY			CAGCCGAGCGCGATCATTTCTAAGG	65,5
panB	GCTAACCAGCTATTGAAAGATGCTC	54,9	GAATATACAGCTTAATGGCAGCACG	56,3
panC	ATTGAAACTTTGCCACTGTTACGCC	59,5	ATACTCACGACGACAACATCGGCAC	61,0
pstS	TGCGGCAAAAGGTGTAGACTGGAGC	64,2	AATCTGAGTTTTCCAGGCAGCACGG	63,3
pstC	ATAAGCCGACAATCAAAGCACCGAG	61,4	CCACAGGAAAGCCCAACCAAATTTC	62,5
argB	TGGCAGTCTGATGAACGTAAATGCC	61,1	CATTCACTTTAACCACCATGCCGTC	60,4
argH	TCAGGCGGCAGATCAGCGTTTTAAG	64,2	GTTGCTGTTCGGCCGTCGTTAAAAC	63,4

Table II Oligonucleotides used for PCR and RT-PCR analyses

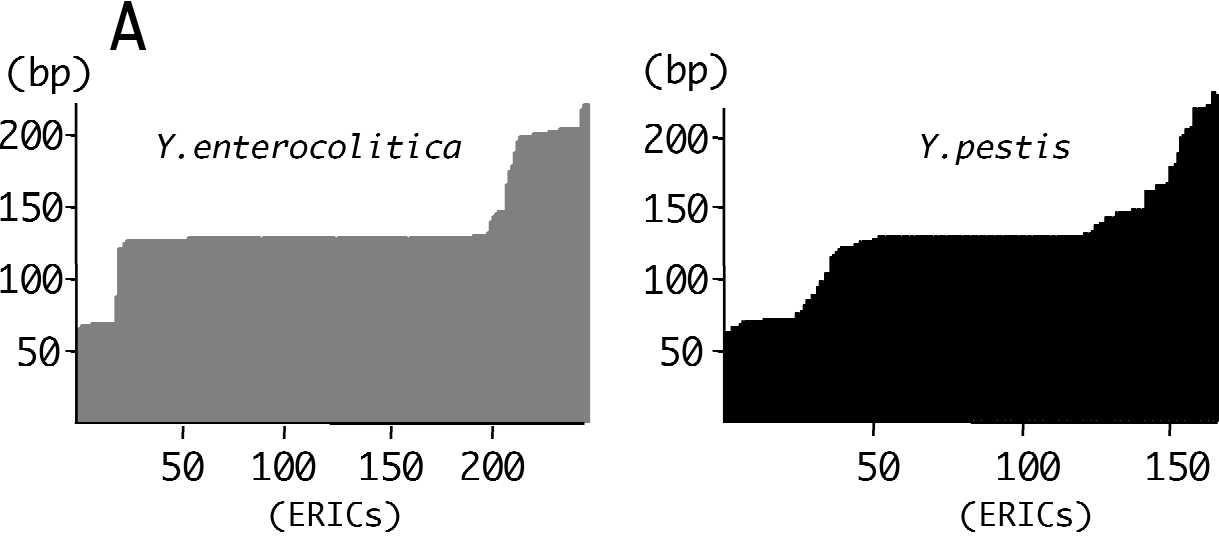
In some instances the level of cistron-specific transcripts was monitored by using novel primers (underlined oligonucleotides). The melting temperature (Tm) of each primer is indicated.

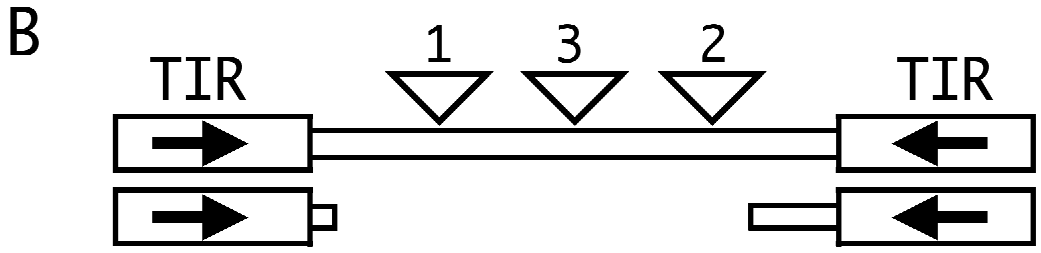
pex.cheW	TTCGTGACGGTTGCTAGTCCTGCCA
pex.trpB	CCCGAACTCGCCAAAATAGGGATTC
pex.uncE	CGATAAAGAACTGTGTACGCAGCAG
pex.lpdA	GGATACAACCGACATTCAGGCACAC

cheB-for	atttaggtgacactatagaaAACTATCAGGTGCGTATTCATGATG
cheB-rev	taatacgactcactatagggGCTTCGTTTTGTGCAATGGTATAAG
cheY-for	atttaggtgacactatagaaTGGTAGACGATTTTTCGACCATGCG
cheY-rev	taatacgactcactatagggTCGGCATGTTCCAGTCAGAAACCAC
panB-for	atttaggtgacactatagaaGCTAACCAGCTATTGAAAGATGCTC
panB-rev	taatacgactcactatagggGAATATACAGCTTAATGGCAGCACG
panC-for	atttaggtgacactatagaaATTGAAACTTTGCCACTGTTACGCC
panC-rev	taatacgactcactatagggATACTCACGACGACAACATCGGCAC

Table III Oligonucleotides used to monitor ERIC+ RNAs

The oligonucleotides used as primers in RNA extension assays are listed at the top. The pairs of 45-mers shown at the bottom have been used to obtain by PCR DNA amplimers in which a copy of the bacteriophage T7 promoter (underlined residues) direct the synthesis of antisense RNA probes (see Fig. 7). Uppercase residues mark *Y. enterocolitica* sequences.





C

TATACCCTAAATAATTCGAGTTGCAGGaaggcg
gcaannngagnga*atcccgatgagctta*caca
agtaagtgattcggg*tganngagngcagccaa
cacaCATGCAGCTTGAAGTATGACGGGTATA

D

type-1 caaatcggtcggnaaccgatttgaacagcatttatgc
tagcccgcagggtgagcctcanngatgaggctcatta

type-2 caaatcngtcgggagcagatttgaacgctgcttgcag
cggcctcgcagaggcgaggcccatggacgggccgagta

type-3 ttattttaaggtcaacg

FIGURE 1

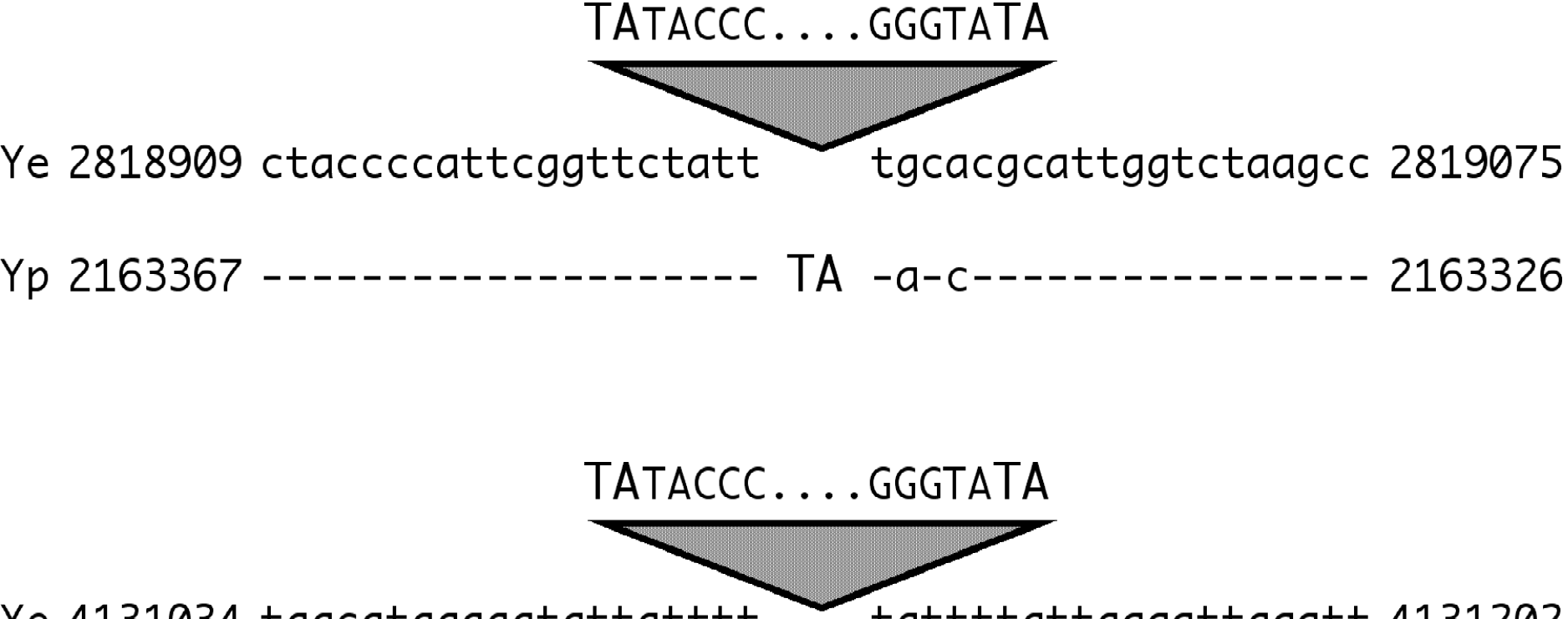

TATACCC....GGGTATA
Ye 2818909 ctaccccattcggttctatt tgcacgcattggtctaagcc 2819075
Yp 2163367 ---------- TA -a-c---------- 2163326
TATACCC....GGGTATA
Ye 4131034 tggcgtgggggtgttgtttt tgttttgttgaaattgaatt 4131202
Yp 4129158 ---------- TA -------c---------- 4129117

FIGURE 2

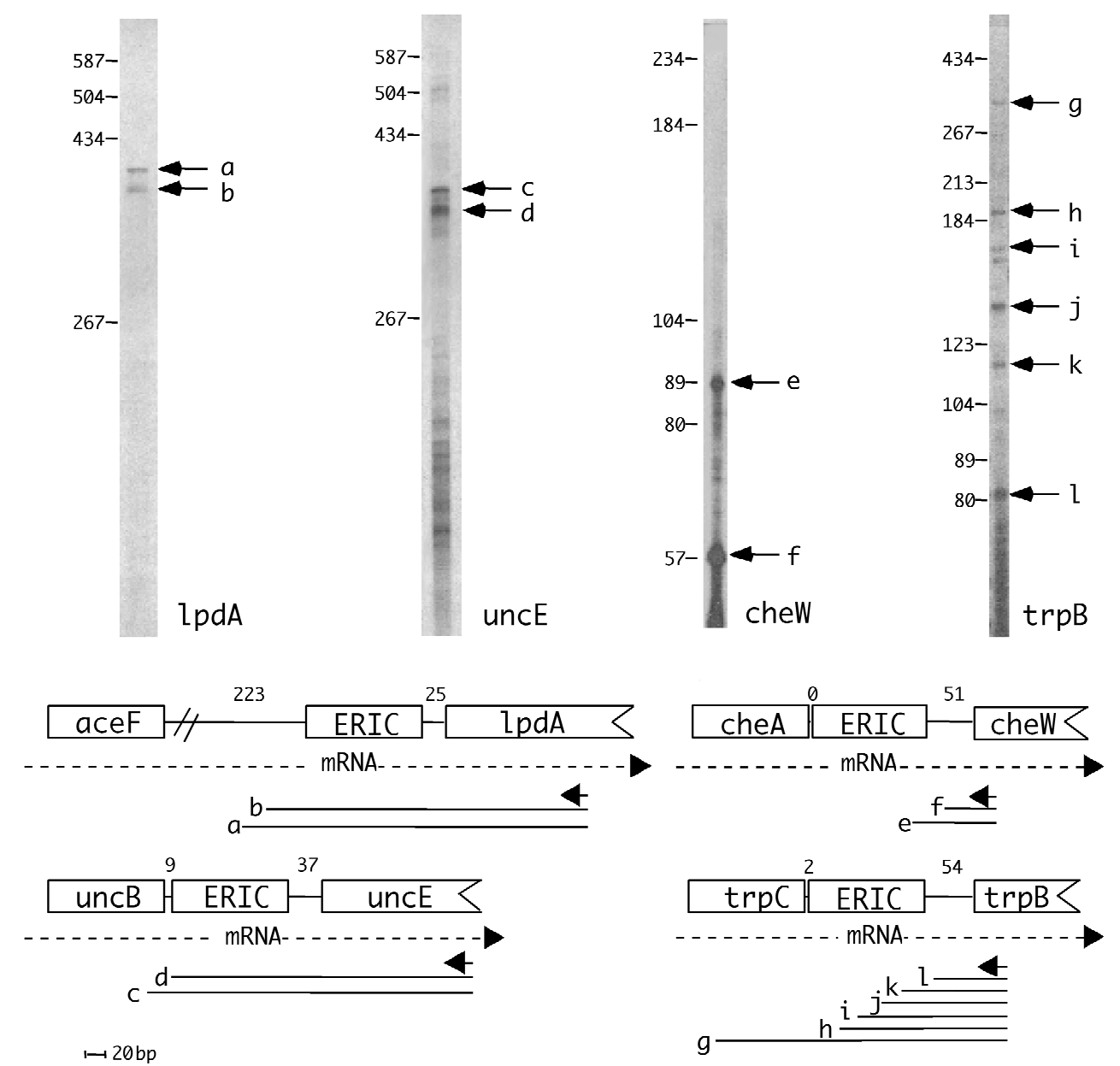

587
504
434
267
a
b
lpdA
587
504
434
267
c
d
uncE
234
184
104
89
80
57
e
f
cheW
434
267
213
184
123
104
89
80
g
h
i
j
k
l
trpB
aceF
223
ERIC
25
lpdA
mRNA
b
a
cheA
0
ERIC
51
cheW
mRNA
f
e
uncB
9
ERIC
37
uncE
mRNA
d
c
trpC
2
ERIC
54
trpB
mRNA
l
k
j
i
h
g
20bp

FIGURE 3

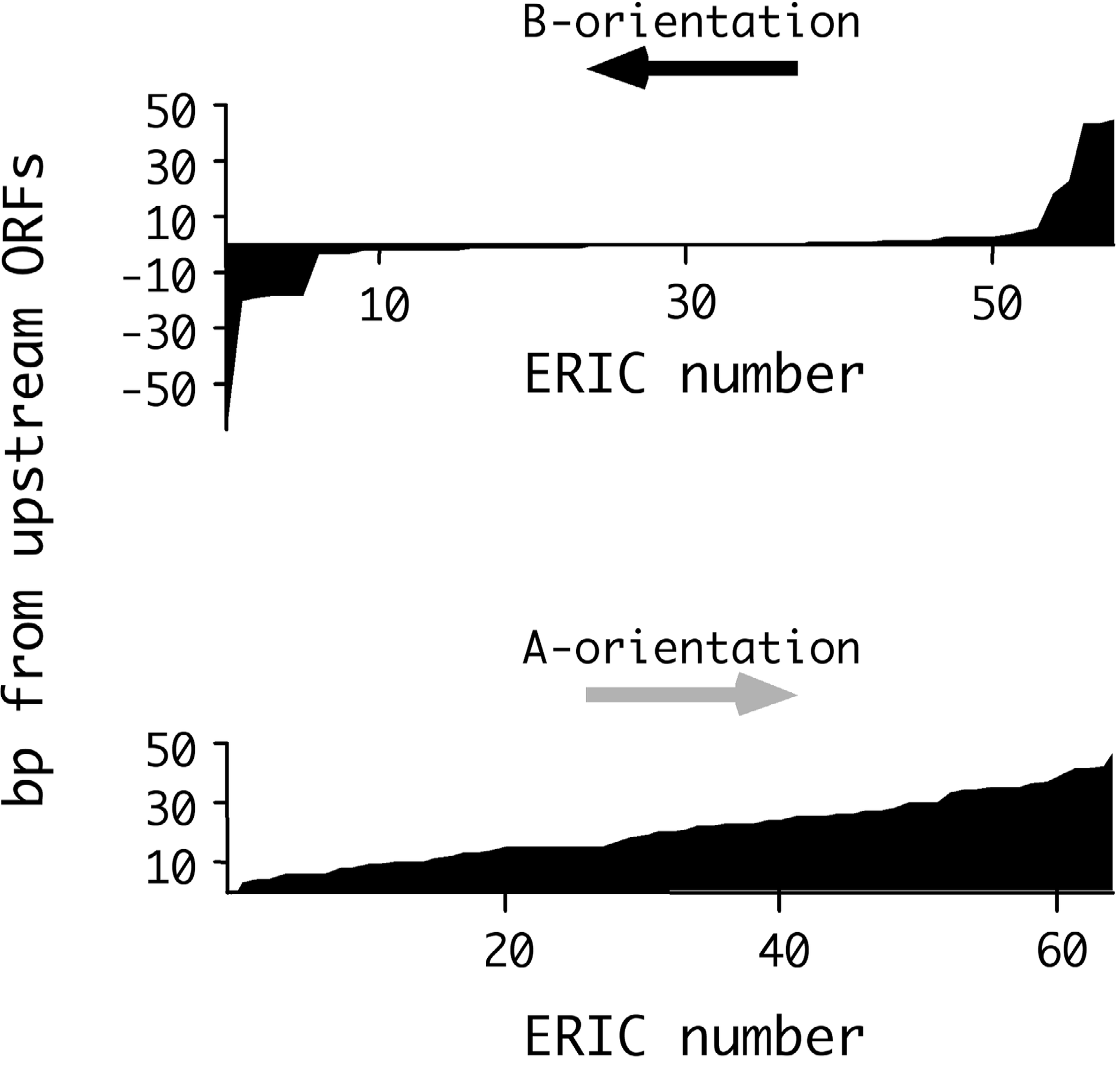

B-orientation
50
30
10
-10
-30
-50
10
30
50
ERIC number
bp from upstream ORFs
A-orientation
50
30
10
20
40
60
ERIC number

FIGURE 4

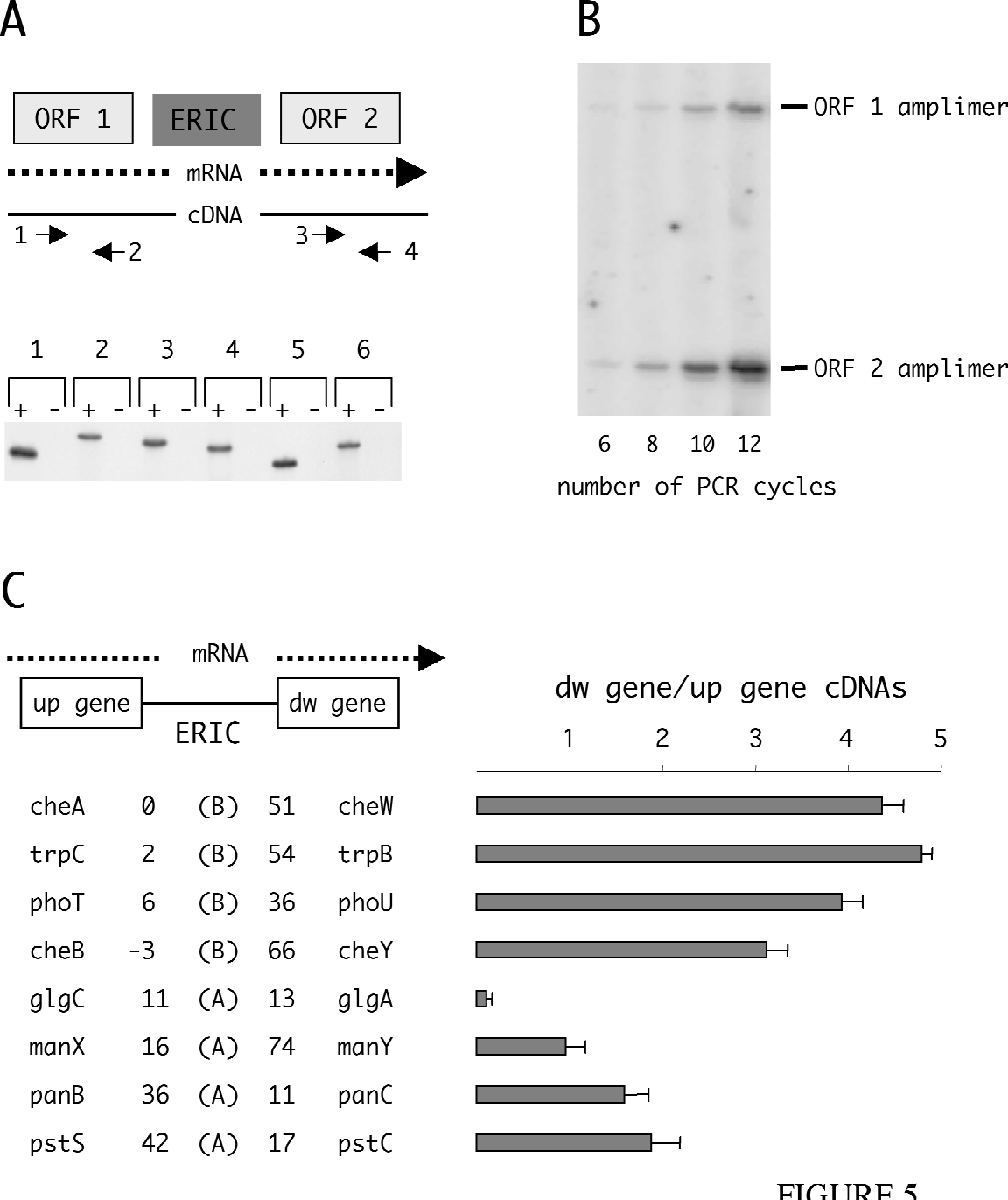

A
ORF 1
ERIC
ORF 2
mRNA
cDNA
1
2
3
4
1 2 3 4 5 6
+ - + - + - + - + - + -
B
ORF 1 amplimer
ORF 2 amplimer
6 8 10 12
number of PCR cycles
C
mRNA
up gene
dw gene
ERIC
dw gene/up gene cDNAs
1 2 3 4 5
cheA 0 (B) 51 cheW
trpC 2 (B) 54 trpB
phoT 6 (B) 36 phoU
cheB -3 (B) 66 cheY
glgC 11 (A) 13 glgA
manX 16 (A) 74 manY
panB 36 (A) 11 panC
pstS 42 (A) 17 pstC

FIGURE 5

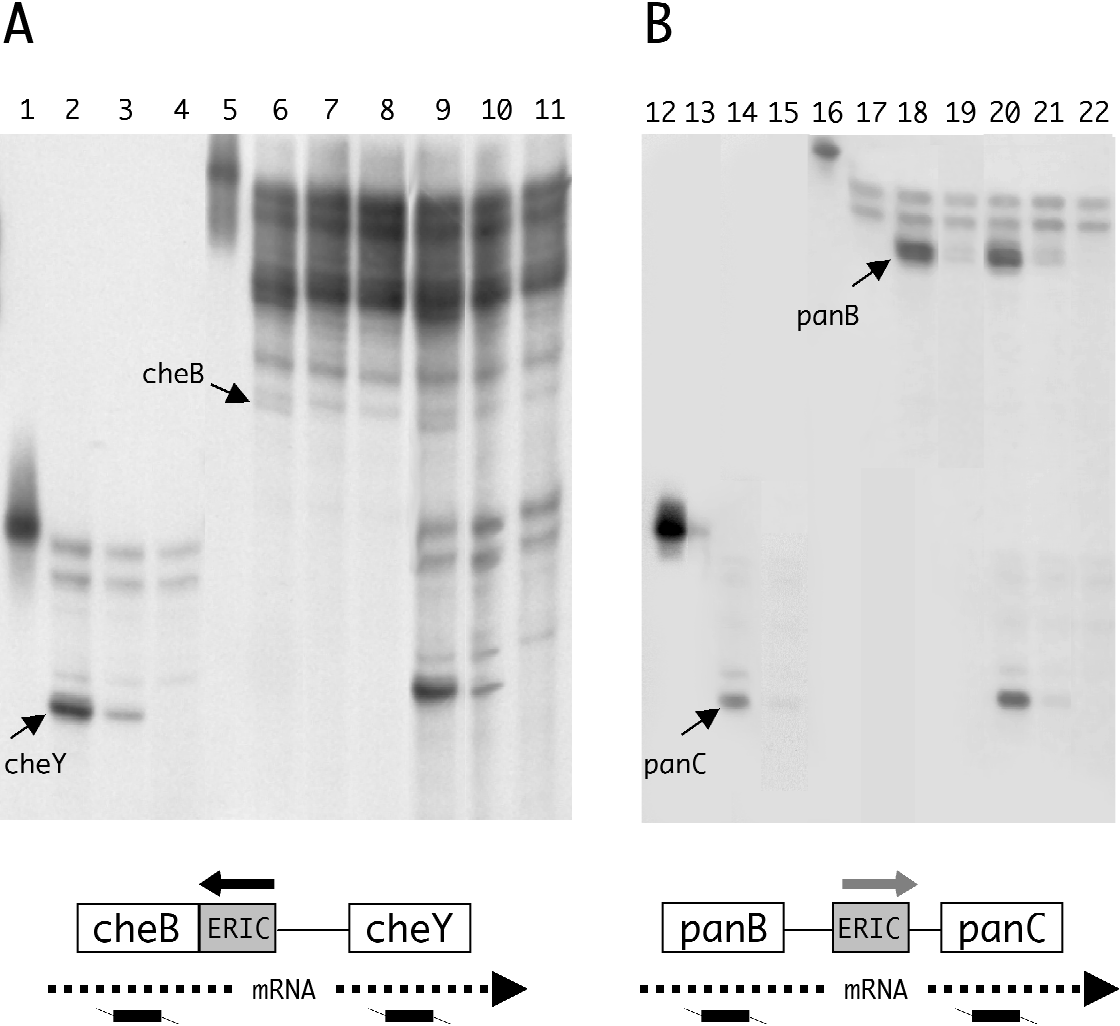

A
1 2 3 4 5 6 7 8 9 10 11
cheB
cheY
B
12 13 14 15 16 17 18 19 20 21 22
panB
panC
cheB
ERIC
cheY
mRNA
panB
ERIC
panC
mRNA

FIGURE 6

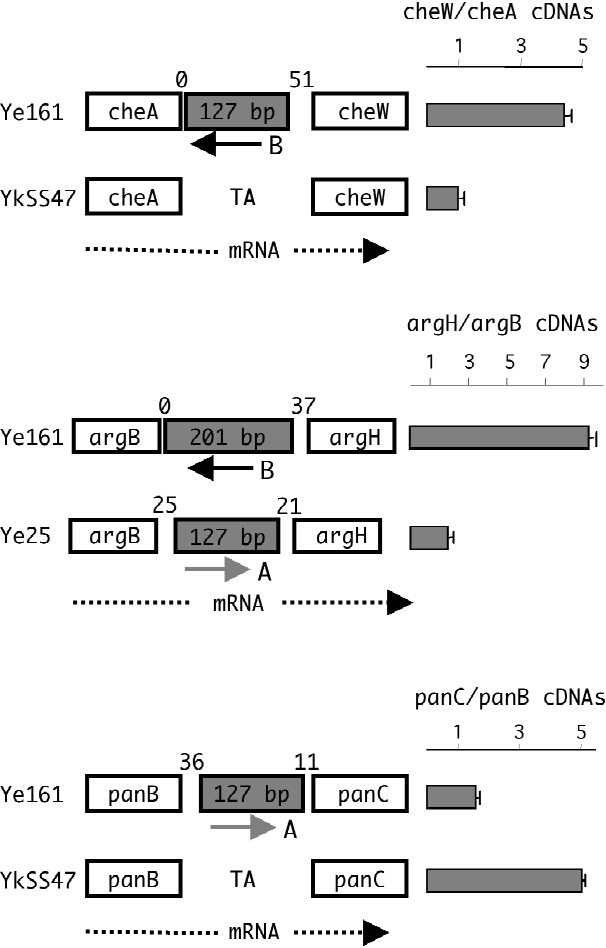

cheW/cheA cDNAs
1 3 5
0
51
Ye161
cheA
127 bp
cheW
B
YkSS47
cheA
TA
cheW
mRNA
argH/argB cDNAs
1 3 5 7 9
0
37
Ye161
argB
201 bp
argH
B
25
21
Ye25
argB
127 bp
argH
A
mRNA
panC/panB cDNAs
1 3 5
36
11
Ye161
panB
127 bp
panC
A
YkSS47
panB
TA
panC
mRNA

FIGURE 7

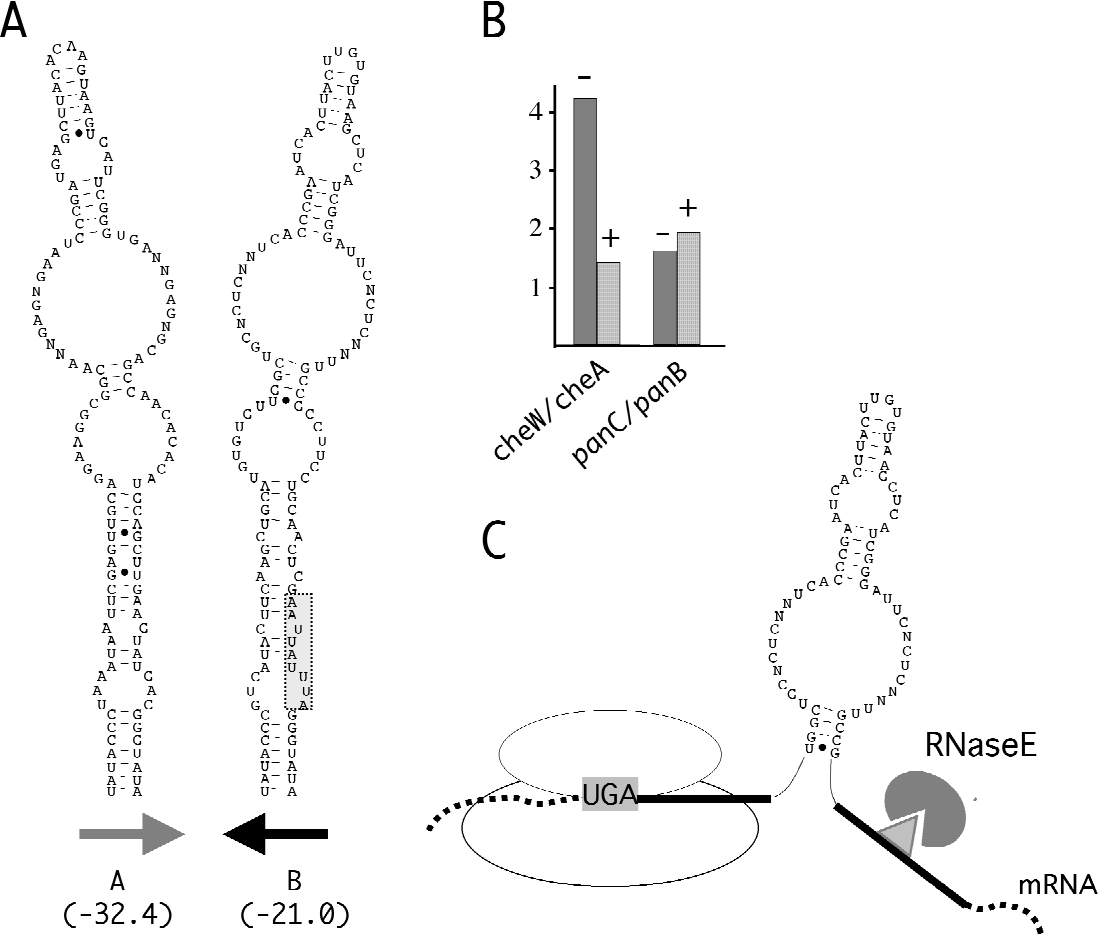


FIGURE 8

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 22, Issue of May 31, pp. 19594–19599, 2002

Printed in U.S.A.

A Novel Intragenic Sequence Enhances Initiator-dependent Transcription in Human Embryonic Kidney 293 Cells*

Received for publication, February 5, 2002, and in revised form, March 25, 2002
Published, JBC Papers in Press, March 26, 2002, DOI 10.1074/jbc.M201193200

Chiara Abrescia‡, Eliana De Gregorio‡, Mattia Frontini§, Roberto Mantovani§, and Pierpaolo Di Nocera‡¶

From the ‡Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli Federico II, Via S. Pansini 5, 80131 Napoli, and the §Dipartimento di Biologia Animale, Università di Modena e Reggio, Via Campi 213/d, 41100 Modena, Italy

In a variety of *Drosophila* TATA-less promoters, transcription is directed by initiator (Inr) sequences, which are faithfully and efficiently recognized only when flanked 3′ by the downstream promoter element (DPE). This motif, which is conserved at ~30 bp from the RNA start site, is viewed as a downstream counterpart to the TATA box, and is recognized by the general transcription factor (TF) IID. By transient expression assays in human embryonic kidney 293 cells, we show that DE1 (distal element 1), a DNA motif located at residues +23 to +29, sustains faithful Inr-dependent transcription as efficiently as the DPE. Transcription significantly increased when DE1 and DPE sequences were adjacently placed on the same template. Results emerging from *in vivo* RNA analyses matched electrophoretic mobility shift assay data. In agarose-electrophoretic mobility shift assays, retarded DNA-protein complexes resulting from the interaction of human holo-TFIID with either Inr^+/DPE^+ or $Inr^+/DE1^+$ promoters were formed at comparable levels, whereas binding of TFIID to both DE1 and DPE motifs was 2-fold increased. The strict requirement for spacing between the Inr and DPE was not observed for DE1, as locating the motif 4 bp away from the +1 site did not impair transcriptional enhancement. DE1 sequences may be common to many promoters and be overlooked because of their poor sequence homology.

A key step in the formation of functional transcription initiation complexes is the recognition of promoter sequences by components of the general transcription machinery. The core promoter sequence context has a significant influence on both the overall efficiency of gene transcription and the ability of individual genes to respond to transcription activators (1). In pol II[1]-dependent transcriptional units, distinct DNA elements have been found to be involved in core promoter function. The TATA box, a sequence located 25–30 bp upstream of the RNA start site, is the key positioning DNA element in many pol II genes (1, 2). The TATA box is recognized by the TATA-binding protein (TBP) subunit of the TFIID complex, a general pol II transcription factor endowed with the ability to recognize promoter DNA (2, 3). In some promoters, the TATA box is immediately preceded by the TFIIB recognition element, fitting the consensus 5′-(G/C)-(G/C)-(G/A)-C-G-C-C-3′. The transcription factor TFIIB plays a central role in preinitiation complex assembly, providing a bridge between promoter-bound TFIID and RNA polymerase II, and TFIIB recognition element increases the affinity of TFIIB for the promoter (4). In many promoters, the TATA element is missing, and is functionally replaced by the initiator (Inr), a stretch of 5–7 residues spanning the RNA start site (5, 6). The Inr is also recognized by TFIID, but physical interactions are mediated by some of the TBP-associated factors, or TAF_{II}s (7, 8). TATA and Inr are functionally exchangeable modules and may coexist in the same gene. The core promoter structure found in a given gene may reflect a preference of the regulators of that gene, and some activators stimulate preferentially TATA-containing or Inr-containing core promoters (9–11).

An additional core promoter module is the downstream promoter element, or DPE. This sequence, conserved ~30 bp downstream from the RNA start site in a variety of *Drosophila* TATA-less promoters, greatly enhances the activity of upstream Inr modules (12–17). DPE interacts with specific components of the *Drosophila* TFIID complex (dTAFII40 and dTAFII60; see Ref. 15). Regions downstream of transcriptional start sites recognized by TFIID, but exhibiting no sequence similarity to DPE, have been identified in a variety of promoters (18–22).

Little is known about the role that intragenic sequences have in promoter recognition and activation in human cells. In this work, we analyzed the transient expression profile of constructs in which Inr sequences are flanked by different types of downstream promoter sequences in human embryonic kidney (HEK 293) cells. Inr-dependent transcription is enhanced by a core DPE sequence located at residues +30/+33. The same holds true for a DNA element called DE1 located at residues +23/+29. DPE and DE1 modules synergize in stimulating transcription *in vivo* and are independently capable, as revealed by agarose-EMSA, to interact with human holo-TFIID.

MATERIALS AND METHODS

Construction of Plasmids—Plasmids described in this work carrying artifical promoters are all derivatives of p8GAL4, a modified pEMBL8CAT vector in which a 54-bp module containing one binding site for the transcriptional activator Gal4 had been inserted at the *Bam*HI site. Sequences homologous to the Inr regions of the *Drosophila* Doc and I LINE promoters (16) have been inserted between the *Bam*HI

* This work was supported by grants from Ministero dell'Università e Ricerca Scientifica, Project "Dinamica della Cromatina nella Espressione Genica" (to P. P. D. N. and R. M.), and Associazione Italiana della Ricerca sul Cancro (to R. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 39-81-7462059; Fax: 39-81-7703285; E-mail: dinocera@cds.unina.it.

[1] The abbreviations used are: pol II, RNA polymerase II; CAT, chloramphenicol acetyltransferase; DCE, downstream core element; DE1, distal element 1; DPE, downstream promoter element; EMSA, electrophoretic mobility shift assay; HEK, human embryonic kidney; hsp70, heat shock protein 70; Inr, initiator; IRF-1, interferon regulatory factor 1; LINE, long interspersed nuclear element; RSV, Rous sarcoma virus; TAF_{II}, TATA-binding protein-associated factor; TBP, TATA-binding protein; TF, transcription factor.

and *Sal*I sites of p8GAL4 to obtain the plasmids G3 and G1, respectively. Constructs G1M, G1X, G1K, G3M, G3X, and G3K, which carry downstream promoter modules, have been obtained by inserting between the *Sal*I and *Hin*dIII sites of either G1 or G3 suitable pairs of complementary oligonucleotides. Plasmids G4M and G5M are derivatives of G3M in which the interval *Bam*HI/*Sal*I spanning the Inr region had been modified. Similarly, the mutant constructs analyzed in Fig. 4 are derivatives of G3M in which the downstream promoter region had been replaced by oligonucleotide pairs having *Sal*I- and *Hin*dIII-compatible termini. To obtain derivatives in which the distance between Inr and downstream sequences increased 4 bp, plasmids of interest were digested with *Sal*I, and reaction products were treated with the Klenow enzyme to fill in gaps prior to ligation and transformation. The control plasmid RSVdel-CAT was obtained by cloning the *Apa*L-*Mlu*I fragment spanning the RSV promoter region in pRSVCAT into the *Hin*dIII site of pEMBL8CAT, and subsequently removing the *Hin*dIII-*Nco*I fragment including most of the CAT coding region. The GAL4-Sp1 plasmid encodes a chimeric Sp1 GAL4 protein containing residues 50–161 of the human Sp1 protein. In all cloning procedures, incompatible termini were blunt-ended by T4 polymerase before ligation. The sequences of the promoter regions analyzed were confirmed by nucleotide sequence analysis.

Cell Culture and DNA Transfections—HEK 293 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Transfection experiments were performed with the standard calcium-phosphate method. Approximately 6×10^5 cells, seeded at a density of 1.2×10^5 cells/ml 24 h prior to transfection, were cotransfected with 10 μg of the plasmid of interest, 0.1 μg of GAL4-Sp1 plasmid, 0.2 μg of RSVdel-CAT plasmid for control of transfection efficiency. Cells were recovered 48 h after transfection, and the activity of constructs was assayed at the RNA level.

RNA Analyses—Total RNA was isolated by using the acid guanidinium thiocyanate/phenol/chloroform single-step extraction method (23). Primer extension assay experiments were performed essentially as described (16). Reaction products were resolved on 8% (w/v) polyacrylamide, 8 M urea gels. Co-electrophoresed sequencing ladders were generated by the dideoxy chain termination method utilizing double-stranded DNA templates. The CAT primer used both to obtain sequencing ladders and to detect transcripts directed by the different promoters constructs has been described previously (24). Transcripts driven by the RSV promoter in the RSVdel-CAT plasmid were detected by using the NCO primer (5′-AGCGGCATCAGCACCTTGTCGCCTTGCGTA-3′), a synthetic 30-mer complementary to a distal interval of the CAT gene sense strand.

Purification of Holo-TFIID and EMSA Analysis—Holo-TFIID was immunopurified from HeLa cells with an anti-TBP antibody as previously detailed (25, 26). EMSAs of TFIID in agarose gels were performed as described in Ref. 27. Three independent preparations of purified TFIID were used in EMSAs. The GAL4-NF-YA fusion protein (28) was obtained by an *in vitro* transcription-translation coupled reticulocyte lysate system (Promega). One μl of GAL4-NF-YA-containing extracts and 10,000 cpm of ^{32}P-labeled fragments were mixed in 10 μl of NF-Y buffer (20 mM Hepes, pH 7.9, 50 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, 1 mM dithiothreitol) and incubated for 20 min at 30 °C. Samples were loaded onto 4.5% polyacrylamide gels (acrylamide/bisacrylamide, 29:1) and electrophoresed in 0.5× TBE buffer. Gels were run at 150 V for 60 min, transferred on no. 3MM paper, and exposed. PCR fragments tested in EMSA analysis span residues −66 to +61 of all promoters but G3, in which the amplified region spans residues −66 to +55. Amplification was obtained by using two synthetic oligomers, the ^{32}P-5′-end-labeled CAT II 30-mer (5′-TCCTTAGCTCCTGAAAATCTCGCCAAGCTT-3′), complementary to the pEMBL8CAT sense strand, and the GT1 54-mer (5′-TCTCGAGCTGCAGCGGAGACTGTCCTCCGAGATCTCTATCACTGATAGGGATCG-3′), homologous to the −66/−12 interval spanning the binding site for the transcriptional activator Gal4 included in each promoter.

RESULTS

Activation of Inr-dependent Transcription in HEK 293 Cells—In several *Drosophila* TATA-less promoters, the DPE promoter element fits the consensus AG(A/T)CGTGY (12, 14). Statistical and biochemical analyses indicate that the 4-bp core DPE sequence G(A/T)CG is sufficient to stimulate Inr transcription (15, 17). In some *Drosophila* long interspersed nuclear element (LINE) promoters, transcription is regulated by complex intragenic regions including DPEs and additional DNA sequence elements (16). To verify whether sequences flanking DPE in LINE promoters could stimulate transcription in mammalian cells, transient expression assays were carried out in HEK 293 cells. In the base plasmids G1 and G3, promoter DNA is uniquely represented by Inr sequences (see Fig. 1). In the other plasmids, Inr sequences are flanked by ~20-bp-long DNA segments containing, at the correct distance, either DE1 or DPE, or both sequences (see Fig. 2). The DE1 sequence is found immediately upstream of a core DPE motif in the *Drosophila* I promoter (16); DPE corresponds to the 4-bp core DPE sequence GACG found in a variety of *Drosophila* promoters (17). In all templates, a GAL4 recognition sequence is centered ~30 bp upstream of the Inr region. Because Sp1 effectively activates Inr^+ promoters (11), each construct was cotransfected with a plasmid encoding the GAL4-Sp1 activator (28). The plasmid RSVdel-CAT was also cotransfected along each construct to provide an internal control.

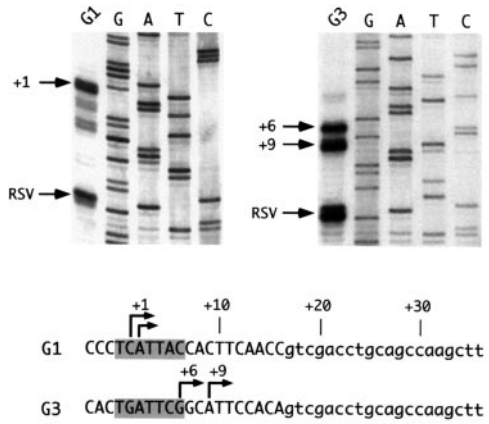


FIG. 1. **Analysis of G1 and G3 promoters.** Total RNA (30 μg) from HEK 293 cells transiently cotransfected with 0.2 μg of the internal reference template RSVdel-CAT plasmid, 0.1 μg of GAL4-Sp1 plasmid, and 10 μg of either G1 (*left panel*) or G3 (*right panel*) plasmid was analyzed by primer extension. Distinct ^{32}P-5′-end-labeled oligomers were used to detect G1-G3 (CAT primer) and RSV (Nco primer) transcripts. Sequencing ladders of the plasmids G1 (*left panel*) and G3 (*right panel*) were obtained by the dideoxy chain termination method using as primer the CAT oligomer. Bands corresponding to major transcripts are marked by *arrows*. The promoter regions of the G1 and G3 templates are aligned at the *bottom*. Inrs (residues −1 to +6) are *highlighted*. *Numbers* are relative to RNA start sites (+1 sites) mapped in *Drosophila* Schneider II cells (16). Vector sequences are in *lowercase letters*.

Correctly initiated transcripts accounted for most of the signal detected with the G1 construct. However, multiple bands marking the accumulation of minor RNA species initiating within the +4 to +8 interval were also detected. By contrast, faithful +1 transcripts driven by the G3 template accumulated at nearly undetectable levels in HEK 293 cells, the prominent signal obtained corresponding to RNAs initiated at residues +6 and +9 (Fig. 1). This peculiar transcriptional pattern plausibly reflects the activity of a secondary Inr module spanning residues +6 to +13 (GGCATTCC; see Fig. 1). In *Drosophila* Schneider II cells, alternative initiation from this secondary Inr is predominant over initiation from the Inr CTGATTC spanning residues −2 to +6 in the absence of DPE (29). The profile of expression of the G3 Inr dramatically changed upon addition of downstream promoter sequences. The presence of either DE1 (G3K construct), DPE (G3X construct), or both (G3M construct) allowed the detection initiation from residue +1 (Fig. 2). Elongation products were quantitated by PhosphorImager analyses, and the efficiency of faithful transcription initiation evaluated as the ratio of transcripts initiated at residues +1 and +6. Quantitative estimates revealed that the $DE1^+DPE^+$ G3M construct directed faithful transcription initiation ~5- and ~10-fold more efficiently than the $DE1^-DPE^+$ G3X and the

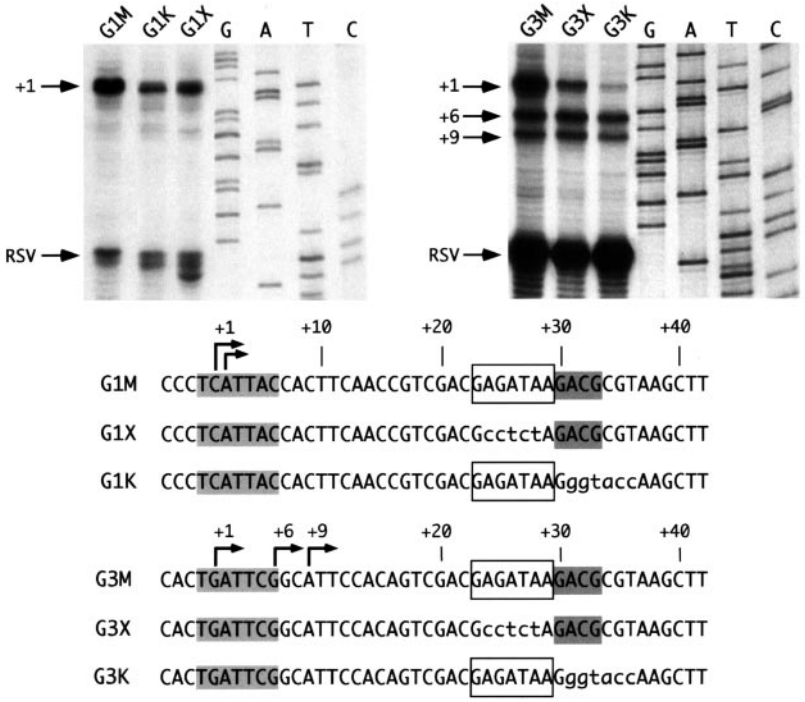


FIG. 2. **Transcriptional reprogramming by DPE and DE1 sequences.** Total RNA (30 μg) from HEK 293 cells transiently cotransfected with 0.2 μg of RSVdel-CAT plasmid, 0.1 μg of GAL4-Sp1 plasmid, and 10 μg of the DNAs indicated at the *top* was analyzed by primer extension as in Fig. 1. Co-electrophoresed sequencing ladders of G1M (*left panel*) and G3M (*right panel*) were obtained by using as primer the CAT oligomer. The promoter regions of the templates assayed are aligned at the *bottom*. Inr, DE1, and DPE sequences are *highlighted*. Base changes altering residues found in G1M and G3M are in *lowercase letters*.

DE1$^+$ DPE$^-$ G3K constructs, respectively.

Thus, DE1 and DPE motifs can reprogram, at comparable levels, the pattern of transcription initiation of the G3 Inr. Moreover, the two motifs synergize in enhancing functional recognition of the +1 site. Similar results were obtained by the analysis of the G1 derivatives G1M, G1K, and G1X (Fig. 2). Comparisons of the autoradiograms shown in Figs. 1 and 2 reveal that initiation at minor sites detected with the parental G1 template was largely inhibited in the three G1 derivatives, each template directing predominantly, if not exclusively, the synthesis of +1 transcripts. To evaluate relative template efficiencies, transcripts directed by the G1 Inr were normalized to transcripts directed by the reference RSVdel-CAT construct. The DE1$^+$DPE$^+$G1M template directed faithful transcription initiation ~2- and ~4-fold more efficiently than the DPE$^+$ G1X and the DE1$^+$ G1K templates, respectively. Differences in the degree of stimulation by the same sequence elements in the G1 and G3 derivatives correlate to differences in the strength of the G1 and G3 Inrs. The G1 Inr, which better fits the optimal Inr consensus (6) and efficiently directs faithful transcription initiation as single module (Fig. 1) is less sensitive to enhancement by downstream promoter sequences. For this reason, we measured the fidelity of initiation as the ratio of transcripts originating from the same template, and subsequent analyses were all carried out with the G3 Inr.

In G3M, both the Inr and the downstream sequences are required to measure efficient transcriptional levels, as shown from the analysis of G4M and G5M, two derivatives in which the *Inr* was variously mutated (Fig. 3). In HEK 293 cells transfected with either plasmid, transcripts initiated at residues +6 and +9 account for most of the detectable signal. Interestingly, transcription initiation at or near residue +1 could still be measured, albeit at very low levels (see bands marked by *gray arrows* in Fig. 3). Heterogeneity in the start sites around position +1 between G4M and G5M may correlate to alternative base selection dictated by the different DNA contexts replacing genuine Inr sequences. These data indicate that, albeit inefficiently, downstream promoter elements contribute to correctly position the transcriptional apparatus even in the absence of a functional Inr.

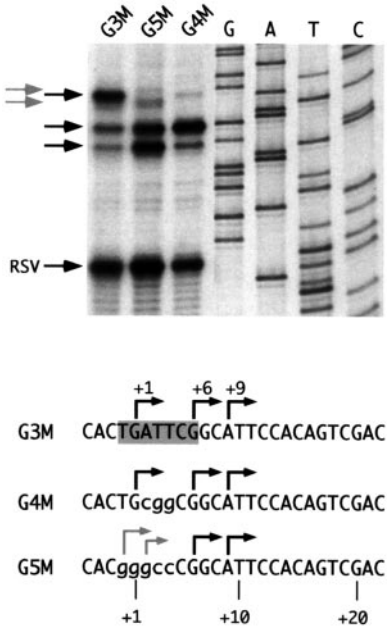


FIG. 3. **Analysis of Inr$^-$ templates.** The template proficiency of G3M, G5M, and G4M in HEK 293 cells was analyzed by primer extension as described in Figs. 1 and 2. Samples were electrophoresed along with a G3M sequence ladder obtained as in Fig. 2. *Black arrows* mark major reaction products, and *gray arrows* minor transcripts driven by the G5M promoter. The G3M, G4M, and G5M start regions are aligned at the *bottom*. Base changes altering the Inr are in *lowercase letters*.

Functional Dissection of the DE1-DPE Region—To characterize the functional interplay between DE1 and DPE, we next transfected HEK 293 cells with derivatives of G3M, in which base changes were introduced within the +23/+33 interval to selectively alter either DNA motif. The base changes and transcriptional proficiencies of the constructs analyzed are reported in Fig. 4. The levels of correct transcription initiation directed by each construct were evaluated by calculating, as in Fig. 2, the +1 *versus* +6 transcript ratio. By looking at templates carrying base changes in the DE1 sequence, mutating positions +25 and +26 (plasmids 34B and 34C) lowered ~2-fold the

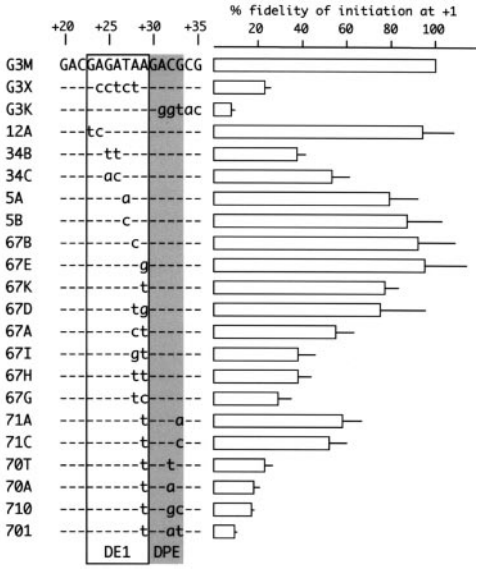


FIG. 4. **Base changes within the downstream promoter region of G3M and transcriptional proficiency.** The downstream promoter regions (residues +20/+35) of the constructs analyzed are aligned to the parental G3M sequence. *Dashes* denote sequence identities. DE1 and DPE motifs are *highlighted*. *Bars* to the *right* denote the efficiency of each template to drive faithful RNA initiation from the +1 site. Data represent the average values obtained in three to five independent transfections. Standard deviations are reported. Values were obtained by calculating first for each template, by PhosphorImager analyses, the +1/+6 transcript ratio, and subsequently by dividing such value by the level of +1/+6 transcripts driven by G3M.

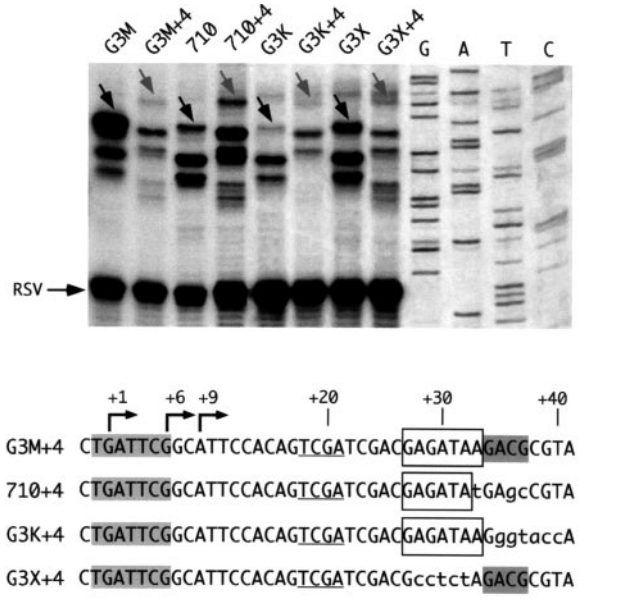


FIG. 5. **Space changes in the promoter region differently influence DPE and DE1 action.** Primer extension analysis of the transcripts directed in HEK 293 cells by the constructs indicated on *top* of the gel, and the G3M sequencing ladder, were obtained as described in the legends to the previous figures. *Black* and *gray arrows* within *lanes* mark +1 transcripts directed by parental and +4 derivative constructs, respectively. The G3M+4, 710+4, G3K+4, and G3X+4 promoter regions are shown at the *bottom*. Inr, DPE, and DE1 sequences are *highlighted* as in Fig. 2. The 4 bp inserted in each construct are *underlined*. Base changes altering residues found in G3M+4 are in *lowercase letters*.

levels of transcription initiation at +1. By contrast, substituting either residues +23 and +24 (construct 12A), or residues +27 (constructs 5A and 5B), +28 (construct 67B), and +29 (constructs 67E and 67K) had no major effect. However, when residues +28 and +29 were both changed, the levels of transcription initiation at +1 dropped 2–3-fold (constructs 67A, 67I, 67H, and 67G), except for the 67D. Thus, crucial residues are located both in the middle of DE1 and at the DE1/DPE boundary. Constructs carrying a mutated DPE motif could be broadly sorted in two main groups. The templates in which the first three DPE residues were preserved (constructs 71A and 71C, Fig. 4), resulted only ~2-fold less efficient than G3M, suggesting that DE1 can still efficiently cooperate with a partial DPE core. When only two adjacent residues of DPE were preserved, functional cooperation between DE1 and DPE was significantly reduced (constructs 710, 70A, 70T, and 701; Fig. 4),, yet three of these constructs, 710, 70A, and 70T, still directed initiation at +1 ~2-fold more efficiently than the DPE^- G3K plasmid. Values are plausibly higher, as in all DPE mutants, the adenine at +29 was replaced by a thymine. The modification, which reduces only slightly the level of +1 transcripts (compare G3M and 67K constructs in Fig. 4), was introduced to enhance the effect of mutations hitting DPE without severely altering the DNA context.

On the whole, our data indicate that correct positioning of the transcriptional pol II machinery could be impaired at comparable levels by mutations affecting either DE1, DPE, or bases between the two motifs. Thus, DE1 and DPE appear to be part of a relatively large DNA region capable of multiple interactions with basal transcriptional factors, and it is therefore not surprising that most of the templates analyzed drive correct transcription initiation efficiently.

DPE, but Not DE1, Functions in a Strict Distance-dependent Fashion—There is a strict requirement for spacing between the Inr and DPE motifs, as an increase, or decrease, of a few nucleotides in the distance between the Inr and DPE causes a severe reduction in transcription. This suggests a specific and somewhat rigid interaction of TFIID with the Inr and DPE sequences (15–17). Interestingly, cooperation between Inr and DE1 is not strictly space-dependent. A pairwise comparison of the template activity of the four constructs G3M, G3X, G3K, and 710 with derivatives in which the distance separating Inr and downstream promoter sequences was increased by 4 bp, is reported in Fig. 5. In all constructs, the efficiency of *Inr*-dependent transcription was quantitated by comparing the levels of transcripts directed by the G3 Inr and the cotransfected RSVdel-CAT construct. Relatively to the parental G3M and G3X DPE^+ templates, faithful transcription initiation was reduced 12-and 4-fold in G3M+4 and G3X+4, respectively. By contrast, G3K and G3K+4, as the templates pair 710 and 710+4, all carrying DE1, directed the synthesis of faithfully initiated transcripts with the same efficiency (Fig. 5). An 8-bp increase in the distance between *Inr* and the downstream region abolished detectable initiation at +1 in all of the templates analyzed (data not shown).

TFIID-Promoter Interactions—To assess whether $DE1^+$ promoters interact with TFIID, the promoter regions of G3, G3M, G5M, G3K, and G3X plasmids were challenged with immunopurified holo-TFIID, and the formation of protein-promoter complexes assessed by electrophoretic mobility shift assays in agarose. In TFIID dose-response experiments, retarded complexes were detected with all the promoter regions assayed (Fig. 6, *panel A*). Each probe contains a GAL4 recognition sequence. The same amount of radiolabeled PCR product was separately incubated with a GAL4-NF-YA fusion protein (28), and retarded GAL4-NFYA/DNA complexes detected on 4.5% polyacrylamide gels (Fig. 6, *panel B*). Quantitative estimates were obtained by PhosphorImager analyses, and relative binding efficiencies were calculated by normalizing probe counts detected in retarded TFIID complexes to probe counts detected in GAL4 retarded complexes (Fig 6, *panel C*). By setting

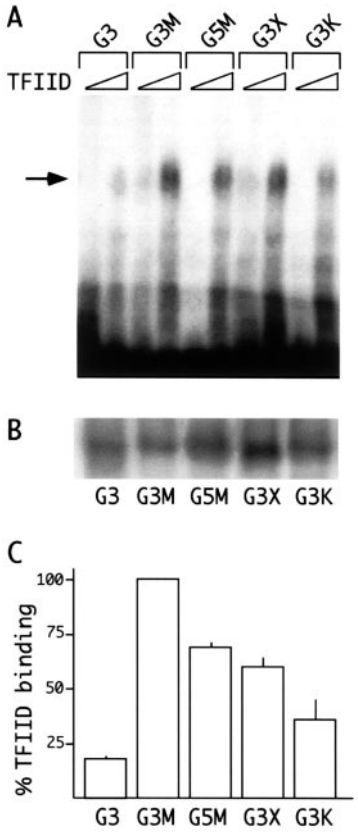


FIG. 6. **Representative agarose-EMSA analyses of TFIID-promoter complexes.** *A*, the ^{32}P-5′-end-labeled CAT II oligomer and the cold GT1 oligomer were used to amplify by PCR the −66/+61 region of the analyzed templates. Approximately 10,000 cpm of each purified PCR product was incubated with either 0.5 or 1.5 μl of an immunopurified human holo-TFIID fraction (see "Materials and Methods"), and samples were loaded onto a 1.2% agarose gel. Bands corresponding to TFIID-DNA retarded complexes are marked by an *arrow*. *B*, the same amount of radiolabeled PCR product was incubated with a GAL4-NF-YA fusion protein. Samples were loaded onto a 4.5% polyacrylamide gel, and GAL4-DNA retarded complexes are shown. *C*, relative TFIID binding efficiencies. Values result from the ratio of TFIID-DNA/GAL4-DNA complexes formed by each template at high TFIID input divided by the TFIID-DNA/GAL4-DNA complex ratio obtained with the G3M probe. Data represent the average values obtained in three to four independent experiments. Standard deviations are reported.

to 100% DNA/TFIID interactions detected with the DE1$^+$DPE$^+$G3M probe, we found that interactions of TFIID were 2-fold less efficient in the G3K template, in which the Inr is flanked by DE1 (36%), and in the G3X template, in which the Inr is flanked by DPE (51%). The efficiency of TFIID binding dropped ~5-fold in the absence of downstream promoter sequences (G3, 18% of binding). In accord to the transfection data shown in Fig. 5, DE1 and DPE are capable to interact with TFIID also in the absence of Inr sequences (G5M probe, 65% of binding).

DISCUSSION

Downstream promoter elements, often found in Inr-dependent promoters, function in part by increasing TFIID-promoter complex formation and/or stability through direct interactions with TAF$_{II}$s. A widely recognized downstream promoter element is DPE, a conserved motif found between residues +28 and +34 in many *Drosophila* transcriptional units. In the fruit fly, this DNA sequence is approximately as common as the TATA box (17). By contrast, inspection of the eukaryotic promoter data base (30) reveals that DPE-like sequences are rarely found, either at the described position or at alternative intragenic windows downstream of the site of RNA initiation, in mammalian promoters. Not surprisingly, DPE modules have been so far identified by functional analyses only in the human TATA-less promoters of the IRF-1 (15) and CD30 receptor (31) genes.

We thought it of interest to examine whether intragenic DNA sequences alternative to DPE, both in terms of sequence content and location relative to the RNA start site, could influence Inr-dependent transcription in a human cellular milieu. DE1 modules flank DPE in some *Drosophila* LINE promoters (16). In this work we showed that, in HEK 293 cells, the DE1 sequence GAGATAA spanning residues +23 to +29 stimulated transcription initiation from upstream Inr sequences nearly as efficiently as a core GACG DPE motif located at residues +30 to +33 (Figs. 2 and 4). Transcriptional enhancement significantly increased when the two motifs were adjacently located on the same template (Figs. 2 and 4). The stimulation, relatively mild on the strong G1 Inr, was magnified when the derivatives of the G3 plasmid were analyzed. In the absence of downstream activating modules, Inr sequences located at the same position of the G1 Inr sequences were not functional in G3 DNA, and transcription initiated preferentially from a secondary Inr (+6 and +9 transcripts, Fig. 1). By contrast, the primary G3 Inr was selectively stimulated by DPE as by DE1 (Fig. 2). Functional interactions between DPE and Inrs located at a specific distance are widely documented (12–16). The finding that DE1 mimics DPE in the activation of the same Inr is novel, and adds knowledge on the range of core promoter elements interacting with the pol II transcriptional machinery.

Results emerging from *in vivo* RNA analyses matched EMSA data, showing that DE1, as well as DPE, increased the stability of the TFIID-DNA complexes (Fig. 6). Qualitatively, the TFIID complexes were not grossly different among the templates used, suggesting that identical, or similar, TBP-containing complexes are involved. Quantitation of agarose-EMSAs revealed that retarded DNA-protein complexes resulting from the interaction with either Inr$^+$/DPE$^+$ or Inr$^+$/DE1$^+$ promoters were formed at comparable levels. TFIID-DNA complexes formed by promoter probes including both DE1 and DPE motifs were 2-fold more abundant (Fig. 6). These results are largely in agreement with the footprinting data previously reported on DPE-containing promoters (14, 15). On the whole, both transfection and biochemical data support the notion that DE1 is an intragenic signal analogous to DPE. Derivatives of the DE1$^+$/DPE$^+$ G3M promoter carrying alterations of DE1 directed faithfully initiated transcripts at least 2-fold more efficiently than the DE1$^-$/DPE$^+$ G3X promoter. Similar results were obtained by analyzing derivatives of the G3M promoter carrying mutated DPE motifs (Fig. 4). The observation that sequence contexts in which either DE1, DPE, or residues at the boundary of the two motifs are changed could sustain transcription with comparable efficiencies, supports the notion that multiple residues within the +23 to +33 region interact with TFIID. A key feature of DPE-driven core promoters is a precise spacing between the Inr and DPE. Noteworthy, functional Inr/DE1 interactions are not as strictly space-dependent as Inr/DPE interactions (Fig. 5). This observation leads us to hypothesize that DE1 and DPE, which cannot be exposed on the same side of the DNA double helix, may interact with different surfaces of TFIID, and possibly contact different TAF$_{II}$s. DPE is bound by TAF$_{II}$60-TAF$_{II}$40 heterotetramers (15). DE1 may stimulate transcription by contacting either TAF$_{II}$250 orTAF$_{II}$150, the two TFIID components involved in recognition of the Inr and sequences further downstream (8, 20), although at present a role in promoter recognition for some of the other TAF$_{II}$s cannot be excluded.

TFIID binds to core promoters through interactions that are

apparently multiple, in that the TATA, Inr, and DPE elements have all been clearly shown to be associated with distinct subunits of the complex (TBP, $TAF_{II}150$-$TAF_{II}250$, and $TAF_{II}60$-$TAF_{II}40$, respectively). This combination of elements serves to maximize TFIID stability on the promoter, thereby contributing to promoter strength (see Ref. 22). It is not, therefore, surprising that both DE1 and DPE may cooperate with Inr to sustain transcription, although the precise rules allowing one element to work in some promoters in the apparent absence of additional contacts are poorly understood. One possibility is that additional factors potentially binding to TFIID, such as TFIIA and NC2, will help in fine-tuning the interactions with core promoter elements, both in a positive and negative way. Intragenic core promoter elements distinct from DPEs have been described in a few promoters. In the human $TATA^+$ megalin/low density lipoprotein receptor-related protein 2 gene, promoter sequences located between positions +5 and +11 (5′-TTTTGGC-3′) interact with TFIID. Downstream contacts do not significantly affect the overall affinity of TFIID binding, but induce dramatic qualitative changes in TFIID interactions in the lipoprotein receptor-related protein 2 TATA box region (21). The human $TATA^+Inr^+$ β-globin promoter contains a large downstream region interacting with TFIID called DCE. Functional DCE subelements map at positions +13/+15, +22/+24, and +31/+33 (see Ref. 22). In the *Drosophila* hsp70 promoter, four regions interact with TFIID: the TATA element, the initiator, and two regions located ~18 and 28 nucleotides downstream of the transcription start site (19). In transgenic flies, Inr and downstream sequences serve overlapping functions, making rather modest contributions to the level of expression of the hsp70 promoter (32). The contributions of individual core sequences could have significant physiological impact in other promoters, and mutations in the β-globin gene DCE subelements II and III are the basis for two kinds of human thalassemia (22). The finding that sequences capable to interact with TFIID found at the same gene coordinates may differently contribute to promoter strength *in vivo* illustrates the difficulties in predicting the functional architecture of core promoters.

Plausibly, DE1-like sequences are not restricted to LINE promoters. This is supported both by statistical and biochemical analyses, indicating that a G nucleotide located 4 bp upstream of the DPE core contributes to transcription from DPE-containing promoters (17). Interestingly, in the G3M promoter, the central G of DE1 is at 4 bp distance from DPE (residue +25, Fig. 2), and is important for DE1 activity (Fig. 4). Notably, a DE1-like motif (5′-GAGGCAA-3′) immediately flanks DPE in the human IRF-1 gene and may account for the residual activity of the IRF-1 promoter upon removal of DPE (15). Finally, a purine-rich sequence partly resembling DE1 (5′-GAGACG-3′) is located at residues +23 to +28 in the middle of the downstream region of the human *gfa* (glial fibrillary acid) promoter, also interacting with TFIID (18). Presumably, DE1 sequences are common to many promoters, but are overlooked because of their poor homology. The consensus resulting from the alignments of three $DE1^+$ *Drosophila* LINE promoters is relatively loose (5′-GRG(A/T)(G/T)AA-3′; see Ref. 16), and different sequences may have DE1 activity, as emerging from the analysis of mutagenized templates in Fig. 4. Sequence flexibility has been similarly observed for DPEs, because the range of sequences that can function as a DPE extend well beyond the GA/TCG motif (17). The analysis of randomized promoter libraries may help to determine the range of functional DE1 sequences and derive position weight matrices used to predict the occurrence of analogous modules in natural promoters as done for TATA and Inr elements (2, 6).

Future analyses may reveal whether DE1 sequences are predominantly found in isolation, or associated to DPE motifs. Transcriptional enhancers that are specific for promoters that contain either DPE or TATA box elements have been elegantly identified by P-mediated transformation analyses in *Drosophila* (33), and it has been shown that the transcriptional repressor NC2 activates DPE-driven promoters and represses TATA-driven promoters *in vitro* (34). In light of these findings, it would be of interest to ascertain whether $DE1^+DPE^+$ and DPE^+ promoters may functionally differ in some of these properties *in vivo*.

Acknowledgment—We thank Valerio Orlando for suggestions and critical reading of the manuscript.

REFERENCES

1. Roeder, R. G. (1996) *Trends Biochem. Sci.* **21,** 327–335
2. Bucher, P. (1990) *J. Mol. Biol.* **212,** 563–578
3. Zawel, L., and Reinberg, D. (1995) *Annu. Rev. Biochem.* **64,** 533–561
4. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998) *Genes Dev.* **12,** 34–44
5. Smale, S. T. (1997) *Biochim. Biophys. Acta* **1351,** 73–88
6. Kraus, R. J., Murray, E. E., Wiley, S. R., Zink, N. M., Loritz, K., Gelembiuk, G. W., and Mertz, J. E. (1996) *Nucleic Acids Res.* **24,** 1531–1539
7. Kaufmann, J., and Smale S. T. (1994) *Genes Dev.* **8,** 821–829
8. Chalkley, G. E., and Verrijzer, C. P. (1999) *EMBO J.* **18,** 4835–4845
9. Kollmar, R., Sukow, K. A., Sponagle, S. K., and Farnham. P. J. (1992) *J. Biol. Chem.* **269,** 2252–2257
10. Colgan, J., and Manley, J. L. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 1955–1959
11. Emami, K. H., Navarre, W. W., and Smale, S. T. (1995) *Mol. Cell. Biol.* **15,** 5906–5916
12. Minchiotti, G., and Di Nocera, P. P. (1991) *Mol. Cell. Biol.* **11,** 5171 5180
13. Contursi, C., Minchiotti, G., and Di Nocera, P. P. (1995) *J. Biol. Chem.* **270,** 26570–26576
14. Burke, T. W., and Kadonaga, J. T. (1996) *Genes Dev.* **10,** 711–724
15. Burke, T. W., and Kadonaga, J. T. (1997) *Genes Dev.* **11,** 3020–3031
16. Minchiotti, G., Contursi, C., and Di Nocera, P. P. (1997) *J. Mol. Biol.* **267,** 37–46
17. Kutach, A. K., and Kadonaga, J. T. (2000) *Mol. Cell. Biol.* **20,** 4754–4764
18. Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R. G., and Freese, E. (1990) *Nature* **348,** 86–88
19. Purnell, B. A., Emanuel, P. A., and Gilmour, D. S. (1994) *Genes Dev.* **8,** 830–842
20. Verrijzer, C. P., Yokomori, K., Chen, J. L., and Tjian, R. (1994) *Science* **264,** 933–941
21. Knutson, A., Castaño, E., Oelgeschläger, T., Roeder, R. G., and Westin, G. (2000) *J. Biol. Chem.* **275,** 14190–14197
22. Lewis, B. A., Kim, T. K., and Orkin, S. H. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7172–7177
23. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
24. Grimaldi, G., and Di Nocera, P. P. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5502–5506
25. Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994) *Cell* **79,** 107–117
26. Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998) *Nature* **393,** 187–191
27. Frontini, M., Imbriano, C., di Silvio, A., Bell, B., Bogni, A., Romier, C., Moras, D., Tora, L., Davidson, I., and Mantovani, R. (2002) *J. Biol. Chem.* **277,** 5841–5848
28. Liberati, C., di Silvio, A., Ottolenghi, S., and Mantovani, R. (1999) *J. Mol. Biol.* **285,** 1441–1455
29. Minchiotti, G., and Di Nocera, P. P. (1997) *FEBS Lett.* **411,** 189–194
30. Perier, R. C., Junier, T., Bonnard, C., and Bucher, P. (1999) *Nucleic Acids Res.* **27,** 307–309
31. Croager, E. J., Gout, A. M., and Abraham, L. J. (2000) *Am. J. Pathol.* **156,** 1723–1731
32. Wu, C. H., Madabusi, L., Nishioka, H., Emanuel, P., Sypes, M., Arkhipova, I., and Gilmour, D. S. (2001) *Mol. Cell. Biol.* **21,** 1593–1602
33. Butler, J. E., and Kadonaga, J. T. (2001) *Genes Dev.* **15,** 2515–2519
34. Willy, P. J., Kobayashi, R., and Kadonaga, J. T. (2000) *Science* **290,** 982–985

INFECTION AND IMMUNITY, July 2003, p. 4217–4221 Vol. 71, No. 7
0019-9567/03/$08.00+0 DOI: 10.1128/IAI.71.7.4217–4221.2003

Asymmetrical Distribution of *Neisseria* Miniature Insertion Sequence DNA Repeats among Pathogenic and Nonpathogenic *Neisseria* Strains

Eliana De Gregorio, Chiara Abrescia, M. Stella Carlomagno, and Pier Paolo Di Nocera*

Dipartimento di Biologia e Patologia Cellulare e Molecolare, "L. Califano," Università degli Studi di Napoli Federico II, 80131 Naples, Italy

Received 31 January 2003/Returned for modification 18 March 2003/Accepted 27 March 2003

***Neisseria* miniature insertion sequences (*nemis*) are miniature DNA insertion sequences found in *Neisseria* species. Out of 57 elements closely flanking cellular genes analyzed by PCR, most were conserved in *Neisseria meningitidis* but not in *N. lactamica* strains. Since mRNAs spanning *nemis* are processed by RNase III at hairpins formed by element termini, gene sets could selectively be regulated in meningococci at the posttranscriptional level.**

DNA repeats known as Correia (4) or *Neisseria* miniature insertion sequences (*nemis* [9]) represent about 2% of *Neisseria meningitidis* genomes (10, 12). These elements mostly differ in the presence and/or absence of a 50-bp long internal segment, contain terminal inverted repeats (TIRs) of variable length (Fig. 1A), and induce the specific duplication of the TA dinucleotide upon genomic integration (3, 8, 9). *nemis* have no coding capacity, and whether they are inactive remnants of larger mobile elements or can still be mobilized by other insertion sequences is unknown.

Intriguingly, most repeats are found inserted close to open reading frames (ORFs). Family members carry transcription initiation (2) and termination (6) signals, and full-length elements contain functional integration host factor sites (3). These observations suggest that *nemis* may impinge on gene expression at the transcriptional level. The finding that *N. meningitidis* mRNAs spanning *nemis* are processed by RNase III at hairpins that are formed by *nemis* TIRs (5, 9) allows one to hypothesize that *nemis* influence the level of expression of neighboring genes mostly by acting at the posttranscriptional level. *nemis* are (or have been) mobile elements, and their distribution in sequenced neisserial genomes is partly different (8, 9). Hence, before concluding on the base of whole-genome data (10, 12) that the expression of specific *N. meningitidis* genes could be regulated by *nemis*-mediated RNase III cleavage, we thought it important to verify the degree of conservation of *nemis* repeats in *N. meningitidis* populations. To this end, the position of a representative set of repeats spread throughout the genomes of the *N. meningitidis* MC58 (Fig. 1B) and Z2491 strains was monitored by PCR analyses in a variety of meningococci and in three strains of the apathogenic species *N. lactamica* (Table 1). The 57 elements selected are inserted close to either the start or the end of neisserial ORFs (Fig. 2). Ten nanograms of DNA from each strain was amplified by using the AmpliTaq DNA polymerase and 100 nanograms of 25- to 30-mers complementary to DNA segments flanking each repeat that were located 300 to 700 bp apart and were designed on the base of sequence conservation among fully sequenced *N. meningitidis* DNAs. Amplimers were resolved by electrophoresis on either 1.4% agarose or 6% polyacrylamide gels, and some were sequenced by the dideoxy chain termination method. In the FAM18 strain, whose sequence is available (http://www.sanger.ac.uk/Projects/N_meningitidis/seroC.shtml), the presence of *nemis* at sites of interest was monitored in silico by BLAST searches (1).

Data are summarized in Fig. 3. Size prediction of the PCR products allowed easy classification of most DNA regions as either "empty" (i.e., lacking *nemis*) or "filled" (i.e., containing *nemis*). Amplimers selected for sequence analysis differed essentially in the presence and/or absence of *nemis* DNA that was replaced in empty sites by TA, the target site duplicated at *nemis* termini. Two major types of variations emerge from our survey. At some sites, long and short *nemis* alternated among *N. meningitidis* strains (see repeats 5, 7, 28, 48, 49, 50, and 55 in Fig. 3). Such heterogeneity likely reflects recombination events that occurred in one strain or a few and eventually spread in neisserial populations by transformation-mediated DNA exchanges. Regions marked by the number sign in Fig. 3 matched neither empty nor filled sites in length and either contained or lacked *nemis* DNA, as shown by Southern and/or sequence analyses. Size identities exhibited by amplimers found in different strains (not shown) suggest that most of these alternative intergenic regions plausibly arose in one strain and were propagated to other clones by transformation.

On the whole, most of the tested repeats were fairly conserved among meningococci belonging to different serogroups and/or sequence types. Thirty-one of 57 elements were found at the same relative position in all the *N. meningitidis* strains analyzed; 11 of 57 were found in all but one or two strains. The degree of conservation of the remaining 15 repeats ranged from 70 to 30%. *nemis* were consistently more conserved in strains belonging to hypervirulent lineages than in other me-

* Corresponding author. Mailing address: Dipartimento di Biologia e Patologia Cellulare e Molecolare, "L. Califano," Universita degli Studi Napoli Federico II, Via S. Pansini 5, 80131 Naples, Italy. Phone: 0039-081-7462059. Fax: 0039-081-7703285. E-mail: dinocera@unina.it.

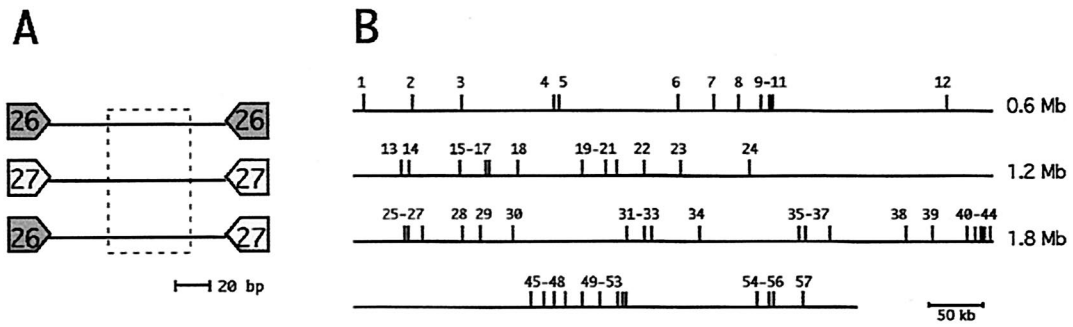


FIG. 1. (A) Organization of *nemis* repeats. *nemis* contain TIRs that are, including the TA dinucleotide target duplicated upon genomic insertion, either 26 or 27 bp. The 50-bp-long central region found only in long elements is boxed. (B) The relative chromosomal positions of *nemis* repeats 1 to 57 used in this study are shown.

ningococci (Fig. 3, bottom panel). Interestingly, the distribution of empty sites among strains is partly lineage specific. Thus, for example, *nemis* 19, 20, and 42 were not found in strains of the ET-5 complex, and *nemis* 19 was also absent in strains of the L1 cluster. *nemis* 55 was absent in lineage 4 strains; *nemis* 51 was absent in strains of both this lineage and the L1 cluster (Fig. 3).

The number of filled sites detected in *N. lactamica* genomes was surprisingly low. Only three repeats were found common to all the strains; 20 were conserved in one to two strains, but 34 were absent from all strains (Fig. 3). Data suggest that *nemis* may be approximately three times less abundant in *N. lactamica* than in *N. meningitidis.* According to in silico analyses, *nemis* are similarly underrepresented in *N. gonorrhoeae* strain F1090 (9), and it is intriguing that most *N. meningitidis nemis*-positive sites are *nemis*-negative sites in both *N. lactamica* and *N. gonorrhoeae* chromosomes (not shown). This would suggest that *nemis* arose in cells ancestral to the divergence of *Neisseriae* in pathogenic and apathogenic species and subsequently spread in a selective fashion in meningococci only.

Many *N. lactamica* regions, shown by Southern analyses to lack *nemis* DNA, are marked by the number sign. These regions not only differed in size from empty sites but varied also in length among strains (not shown) and represent either vestiges of *nemis*-positive intervals or never experienced the insertion of *nemis*. In either instance, it is intriguing that, while genes analyzed occupy the same position in *N. meningitidis* and *N. lactamica* and hence were detected by PCR, the corresponding intergenic regions evolved differently in the two species.

Taking into account that DNA exchanges between pathogenic and apathogenic *Neisseria* species are plausibly as fre-

TABLE 1. Strains used in this study

Species and strain	Serogroup	Epidemiological group	Origin	Source[a]
N. meningitidis				
BF2	B	ET-37 complex	Italy	a
93/4286	C	ET-37 complex	Norway	b
NGP165	B	ET-37 complex	Norway	b
FAM18	C	ET-37 complex	United States	World Wide Web
BZ169	B	ET-5 complex	The Netherlands	b
H44/76	B	ET-5 complex	Norway	b
MC58	B	ET-5 complex	Scotland	World Wide Web
205900	A	Subgroup IV-1	Italy	b
Z2491	A	Subgroup IV-1	The Gambia	World Wide Web
BL859	B	Lineage 3	Italy	c
BS845	B	Lineage 3	Italy	c
BL892	B	Lineage 3	France	d
BF9	B		Italy	a
B1940	B		Germany	e
BL947	B		France	d
NGF26	B		Norway	b
NGE31	B		Norway	b
NGH15	B		Norway	b
N. lactamica				
21			France	d
411			France	d
4627			France	d

[a] a, II Policlinico, Università di Napoli, Naples, Italy; b, IRIS, Chiron S.p.A, Siena, Italy; c, Istituto Superiore di Sanità, Rome, Italy; d, Institut Pasteur, Paris, France; and e, Bayerische Julius-Maximilians Universität, Würzburg, Germany.

ORF	gene function		*nemis*		gene function	ORF
11	murA, UDP-N-acetylgl. carboxyvinyltransf.	26 u	1	d 310	transmembrane transport protein	12
51	pilU, twitching motility protein	100 d	2	u 30	integral membrane protein	50
89	pykA, pyruvate kinase	173 d	3	u 34	outer membrane protein	88
189	integral membrane protein	33 u	4	u 222	conserved hypothetical protein	188
194	alanine symporter protein	135 d	5	u 35	gidA, regulatory protein	193
295	ffh, signal recognition particle protein	62 d	6	d 114	dsbA, thiol:disulphide interchange protein	294
329	pilF, type IV assembly protein	46 u	7	d 289	conserved hypothetical protein	328
353	hypothetical protein	67 u	8	d -4	hypothetical protein	352
380	anaerobic transcriptional regulator	35 u	9	u 122	hemN, coproporphyrinogen III oxidase	379
388	integral membrane transport protein	28 u	10	u 204	ABC transporter ATP-binding protein	387
390	mapA, maltose phosphorylase	45 u	11	d 35	galM, aldose 1-epimerase	389
534	membrane protein	45 u	12	u 390	transmembrane hexose transporter	535
614	putative oxidoreductase	13 d	13	d 76	amtB, probable ammonium transporter	615
619	conserved hypothetical protein	62 d	14	d 21	phosphoglycolate phosphatase	620
671	putative malate oxidoreductase	63 d	15	u 330	tetraacyldisaccharide 4'-kinase	672
698	unknown protein	40 d	16	u 40	tryptophan synthase, beta subunit	699
699	tryptophan synthase, beta subunit	14 d	17	d 59	IgA endopeptidase	700
723	rplT, 50S ribosomal protein L20	114 d	18	u 125	pheS, phenylalanyl-tRNA synthetase a chain	724
785	recB, exodeoxyribonuclease V beta chain	28 u	19	d 50	unknown protein	786
811	murB, UDP-N-acetylenolpyruvoylgl. reduct.	53 u	20	d 19	transmembrane efflux protein	812
823	adk, adenylate kinase	320 d	21	u 49	pyrF, orotidine 5'-phosphate decarboxylase	824
845	PhoH-related protein	39 u	22	d -57	LPS biosynthesis related protein	846
885	dnaB, putative replicative DNA helicase	126 d	23	u 26	FimT, fimbrial protein	886
956	sucB, dihydrolipoamide succinyltransf. E2 comp.	49 d	24	u 209	lpdA, dihydrolipoamide dehydrogen. E3 comp	957
1241	cca, tRNA nucleotidyltransferase	21 d	25	u 96	hypothetical protein	1242
1244	rpe, ribulose phosphate epimerase	23 u	26	d 184	hypothetical protein	1245
1257	site-specific DNA methylase, pseudogene	18 u	27	d 123	unknown protein	1258
1287	iron sulphur binding protein	17 u	28	d 101	nrdB, ribonucleoside-diphosphate reductase	1288
1302	hip, integration factor beta subunit	112 d	29	d 84	putative transcriptional regulator	1303
1331	uvrB, excinuclease ABC subunit B	124 u	30	d 99	prc, carboxy-terminal processing protease	1332
1421	nifR3 protein	150 u	31	u 62	ATP-dependent RNA helicase	1422
1433	putative lipoprotein	50 u	32	d 3	phospholipase D family protein	1434
1438	hypothetical iron-sulphur protein	57 d	33	d 19	purE, phosphoribosylaminoimidazole carboxyl.	1439
1474	possible tautomerase	90 u	34	u 99	possible periplasmic protein	1475
1558	dgk, diacylglycerol kinase	25 u	35	d 132	gshB, glutathione synthetase	1559
1563	transcriptional regulator [GntR-family]	102 u	36	d 315	hypothetical protein	1564
1584	unknown protein	88 d	37	u 34	transcriptional regulator [MarR-family]	1585
1650	transcriptional regulator [AsnC-family]	43 u	38	u 174	alr, alanine racemase	1651
1669	hemO, haem utilisation protein	62 u	39	u 66	integral membrane protein	1670
1695	hypothetical protein	-89 d	40	d 139	hypothetical protein	1699
1706	unknown protein	-5 d	41	u 27	integral membrane ion transporter	1707
1710	gdhA, glutamate dehydrogenase	96 d	42	d 38	transcriptional regulator [GntR family]	1711
1711	lrp, transcriptional regulator [GntR family]	42 u	43	u 234	integral membrane protein	1712
1716	mtrC, membrane fusion protein	57 u	44	u 192	transcriptional regulator [mtrR family]	1717
1861	prmA, ribosomal protein L11	18 u	45	d 62	accC, acetyl-CoA carboxylase	1862
1877	prolyl endopeptidase	28 u	46	d 65	argA, acetylglutamate synthase	1876
1883	hypothetical protein	101 d	47	d 30	ferric siderophore receptor protein	1882
1897	leuS, leucyl-tRNA synthetase	159 d	48	u 23	drg, type II restriction endonuclease	1896
1918	fabD, malonyl CoA-acyl c. p. transacylase	74 u	49	d -44	integral membrane protein	1917
1933	atpC, ATP synthase epsilon chain	146 d	50	u 103	glyQ, glycyl-tRNA synthetase alpha chain	1932
1953	sspA, stringent starvation protein A	42 u	51	d 86	hypothetical protein	1954
1956	rpmE, 50S ribosomal protein L31	81 d	52	u 66	cad, cadmium resistance protein	1955
1957	putative acetyltransferase	21 u	53	u 109	rpmE, 50S ribosomal protein L31	1956
2056	rpsI, 30S ribosomal protein S9	105 d	54	d 43	transcriptional regulator [metR family]	2055
2066	tldD, regulatory function	46 u	55	u 237	conserved hypothetical protein	2067
2071	thiG, thiamine biosynthesis protein	14 u	56	d 43	unknown protein	2070
2104	mafA, adhesin	45 u	57	d 61	pyrH, uridylate kinase	2103

FIG. 2. *nemis* were analyzed. Flanking ORFs are numbered as in the *N. meningitidis* MC58 strain (12). The distance in base pairs separating *nemis* from upstream (u) and downstream (d) ORFs is given.

quent as those occurring between meningococci (7), the asymmetry in the partition of *nemis*-positive and *nemis*-negative intergenic regions between *N. meningitidis* and *N. lactamica* strains is striking. This permits the hypothesis that the persistence of *nemis* DNA at specific chromosomal sites may be functional to meningococci.

Many *N. meningitidis* genes listed in Fig. 2 have a functional role. Some encode either transcriptional regulators (ORFs 380, 1585, 1650, and 1711) or regulatory proteins (ORFs 193, 1953, and 2066); others encode proteins known to be involved in pathogenesis (ORFs 329, 700, and 886) or shown to be essential for the development of bacteremia in the rat (ORFs 1422, 1558, and 1671) (12). Transcripts spanning the underlined ORFs are processed at *nemis* RNA hairpins (5, 9). The same holds for mRNAs spanning the additional ORFs listed in Fig. 2 (unpublished results). The hypothesis that *nemis*-mediated RNA processing may have relevance in the life of meningococci as pathogens is strengthened by the observation that RNase III, while dispensable for viability, is crucial for the survival of meningococci in the infected host (11).

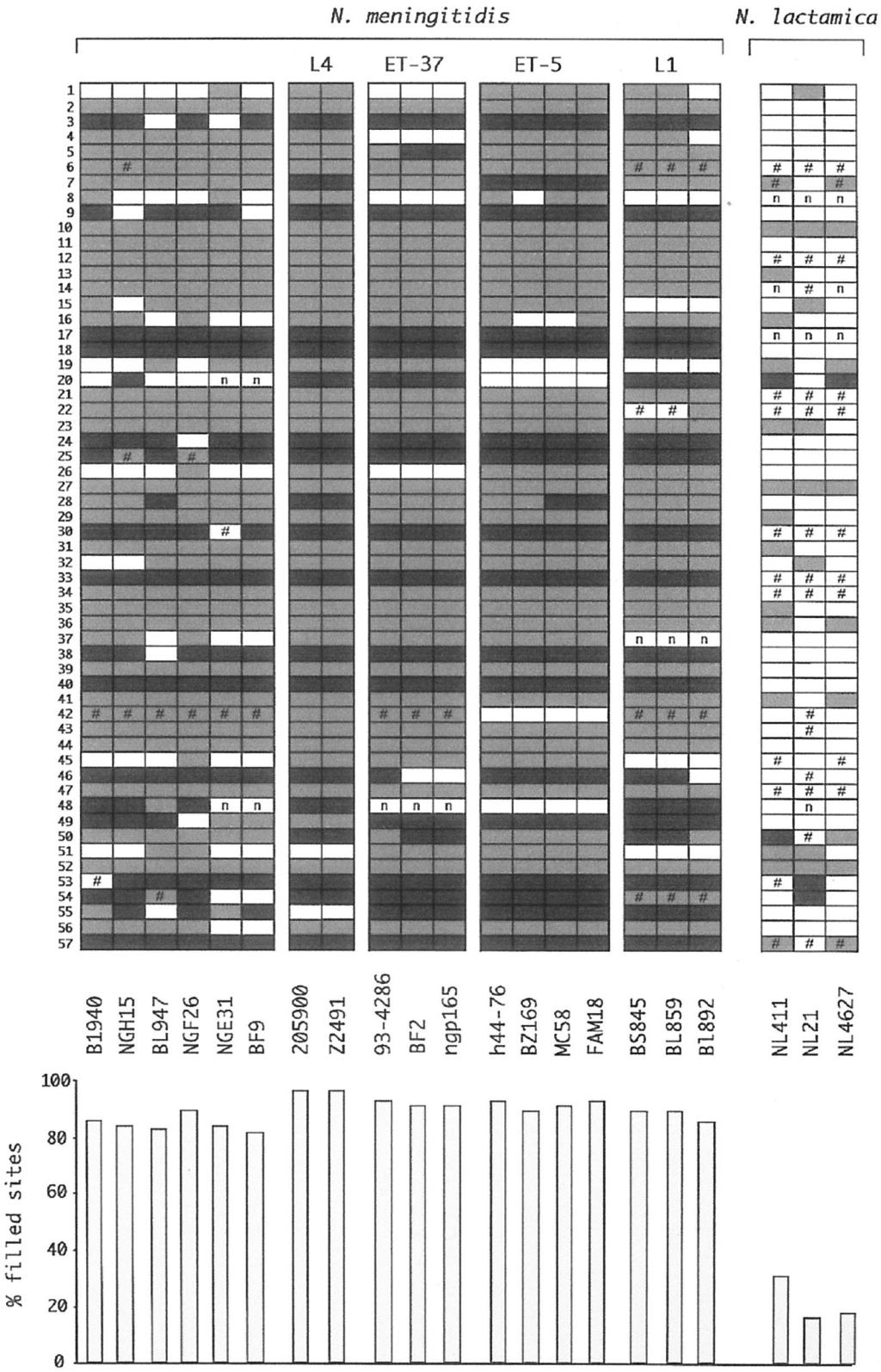


FIG. 3. Conservation of *nemis* repeats in neisserial chromosomes. The distribution of the 57 *nemis* elements listed in Fig. 2 is diagrammed as follows: empty and filled boxes represent intergenic chromosomal regions lacking and containing *nemis* elements, respectively. The presence of long and short *nemis* is marked by light and dark grey filling, respectively. The number sign represents regions differing in size from either filled or empty sites. Regions for which reliable PCR amplification signals could not be obtained, regardless of changes in either PCR settings or primer pairs, are labeled by n. The relative abundance of *nemis*-positive regions within each strain is highlighted in the histogram at the bottom.

We thank Caterina Pagliarulo for providing us with neisserial strains and Giustina Silvestro for help in biocomputing analyses.

This work was partly supported by a grant of program PRIN 2002 of MIUR to P.P.D.N.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Black, C. G., J. A. Fyfe, and J. K. Davies.** 1995. A promoter associated with the neisserial repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. J. Bacteriol. **177:**1952–1958.
3. **Buisine, N., C. M. Tang, and R. Chalmers.** 2002. Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic Neisseria species. FEBS Lett. **522:**52–58.
4. **Correia, F. F., S. Inouye, and M. Inouye.** 1988. A family of small repeated elements with some transposon-like properties in the genome of Neisseria gonorrhoeae. J. Biol. Chem. **263:**12194–12198.
5. **De Gregorio, E., C. Abrescia, M. S. Carlomagno, and P. P. Di Nocera.** 2002. The abundant class of *nemis* repeats provides RNA substrates for ribonuclease III in Neisseriae. Biochim. Biophys. Acta **1576:**39–44.
6. **Francis, F., S. Ramirez-Arcos, H. Salimnia, C. Victor, and J. R. Dillon.** 2000. Organization and transcription of the division cell wall (dcw) cluster in Neisseria gonorrhoeae. Gene **251:**141–151.
7. **Linz, B., M. Schenker, P. Zhu, and M. Achtman.** 2000. Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*. Mol. Microbiol. **36:**1049–1058.
8. **Liu, S. V., N. J. Saunders, A. Jeffries, and R. F. Rest.** 2002. Genome analysis and strain comparison of Correia repeats and Correia repeat-enclosed elements in pathogenic *Neisseria*. J. Bacteriol. **184:**6163–6173.
9. **Mazzone, M., E. De Gregorio, A. Lavitola, C. Pagliarulo, P. Alifano, and P. P. Di Nocera.** 2001. Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic Neisseriae. Gene **278:**211–222.
10. **Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, et al.** 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature **404:**502–506.
11. **Sun, Y. H., S. Bakshi, R. Chalmers, and C. M. Tang.** 2000. Functional genomics of Neisseria meningitidis pathogenesis. Nat. Med. **6:**1269–1273.
12. **Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, et al.** 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287:**1809–1815.

Editor: J. N. Weiser

Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs

Eliana DE GREGORIO, Chiara ABRESCIA, M. Stella CARLOMAGNO and Pier Paolo DI NOCERA[1]

Dipartimento di Biologia e Patologia Cellulare e Molecolare "L. Califano", Università degli Studi di Napoli Federico II, Via S. Pansini 5, 80131 Napoli, Italy

Approx. 2 % of the *Neisseria meningitidis* genome consists of small DNA insertion sequences known as Correia or *nemis* elements, which feature TIRs (terminal inverted repeats) of 26–27 bp in length. Elements interspersed with coding regions are co-transcribed with flanking genes into mRNAs, processed at double-stranded RNA structures formed by TIRs. *N. meningitidis* RNase III (endoribonuclease III) is sufficient to process *nemis*$^+$ RNAs. RNA hairpins formed by *nemis* with the same termini (26/26 and 27/27 repeats) are cleaved. By contrast, bulged hairpins formed by 26/27 repeats inhibit cleavage, both *in vitro* and *in vivo*. In electrophoretic mobility shift assays, all hairpin types formed similar retarded complexes upon incubation with RNase III. The levels of corresponding *nemis*$^+$ and *nemis*$^-$ mRNAs, and the relative stabilities of RNA segments processed from *nemis*$^+$ transcripts *in vitro*, may both vary significantly.

Key words: DNA insertion sequence, pathogenic micro-organism, post-transcriptional control, RNA hairpin, segmental RNA stability.

INTRODUCTION

Neisseria meningitidis (the meningococcus) is a strictly human Gram-negative micro-organism that may either be harmless or cause life-threatening infections, especially in young adults. Meningococci colonize the nasopharynx, but may invade the epithelium, disseminate within the bloodstream and cross the blood–brain barrier, causing septicaemia and meningitis. *N. meningitidis* are classified into a few major pathogenic serogroups. In all cases, a restricted number of clonal groups, or sequence types, cause pathology [1–3]. The determination of the complete genomic sequences of the *N. meningitidis* A serogroup Z2491 [4] and B serogroup MC58 [5] strains have provided data relevant to the molecular characterization of meningococci [6,7]. In parallel, *N. meningitidis* genes essential for the development of bacteraemic disease were identified by functional screening of signature-tagged mutants *in vivo* [8].

Whole-sequence data have also provided clues about the unique organization of Neisserial chromosomes. *N. meningitidis* genomes host an unexpected variety of intergenic DNA repeats, which vary in length and relative abundance, and are found both in isolation and in arrays of 200–2700 bp [4]. *Neisseriae* are micro-organisms that are naturally competent for DNA transformation, and their genome undergoes constant remodelling [9–11]. It is plausible to hypothesize that intergenic DNA repeats may promote recombination by increasing the rate of horizontal gene transfer at flanking loci [4].

A major class of DNA repeats found in Neisserial genomes is represented by short ISs (insertion sequences) which feature 26–27 bp long TIRs (terminal inverted repeats), known as Correia elements [12] or *nemis* (for neisseria miniature insertion sequences; [13]). The genomic insertion of these repeats is accompanied by the duplication of the TA dinucleotide [13–15].

Nemis account for ≈ 2 % of the genome in *N. meningitidis* strains that have been completely sequenced [4,5]. Approx. one-third of the elements are intermingled with repeated DNA sequences in intergenic arrays. The remaining members of the *nemis* family are interspersed with coding regions, and are frequently found inserted close to either the start or the end of ORFs (open reading frames) [13,15]. This unusual distribution, which is fairly well conserved among *N. meningitidis* strains of different serogroups, sequence types and geographical origins [16], allows one to hypothesize that *nemis* sequences have a role in the control of gene expression. It has been reported that individual repeats may promote transcription initiation and termination in *Escherichia coli* and *N. gonorrhoeae* respectively [17,18], and a functional site for the transcriptional activator IHF (integration host factor) [19] has been identified in full-length *nemis* copies [14]. From these observations, it could be argued that *nemis* may impinge on the expression of flanking genes by acting variably at the transcriptional level. In contrast, the finding that *N. meningitidis* mRNAs spanning *nemis* are cleaved at hairpins formed by *nemis* TIRs [13,20] leads one to hypothesize that several members of the *nemis* family influence the expression of neighbouring genes by acting at the post-transcriptional level.

In the present paper we show that the *N. meningitidis* RNase III (endoribonuclease III) expressed *in vitro* is sufficient to process *nemis* RNA hairpins. Transcripts spanning *nemis* may be either sensitive or refractory to processing, and comparative analyses show that the segmental stability of processed *nemis*$^+$ transcripts may vary significantly *in vitro*, as *in vivo*.

EXPERIMENTAL

Bacterial strains and extracts

N. meningitidis and *N. lactamica* strains have been described previously [16,20], and were grown in GC broth supplemented with 1 % Polyvitox (Bio-Merieux). The preparation of whole-cell extracts from *E. coli* and *N. lactamica* has been described in [20].

Expression of *N. meningitidis* RNase III

To clone the *N. meningitidis rnc* gene encoding RNase III, two 31-mers (RNase up, 5′-GAAAGTTGCTGCAGACGATGTTTT-GAAACAG-3′; RNase down, 5′-CGAATCAAGCTTGCCGCC-TCATTTCTTTTTC-3′; underlined residues were deduced from

Abbreviations used: IS, insertion sequence; ORF, open reading frame; REP, repetitive extragenic palindrome; TIR, terminal inverted repeat.

[1] To whom correspondence should be addressed (e-mail dinocera@unina.it).

Table 1 Oligonucleotides used to construct the mini-*nemis* templates

All oligonucleotides are shown in the 5′ to 3′ direction. Residues corresponding to *nemis* TIRs are underlined.

Oligonucleotide	Sequence
Mini-26/26a	GGGAGACAAGAATAAACGCTGAACTATAGTGGATTAACAAAAATCAGGACAAG
	ATGCGCCGTACTGGTTTTTGTTAATCCACTATACACGAACTCAACGCTGTAAC
Mini-26/27	GGGAGACAAGAATAAACGCTGAACTATAGTGGATTAACAAAAATCAGGACAAG
	ATGCGCCGTACTGGTTTAAATTTAATCCACTATACACGAACTCAACGCTGTAAC
Mini-26/26b	GGGAGACAAGAATAAACGCTGAACTGTAAATGAAGTACAAAAATCAGGACAAG
	ATGCGCCGTACTGGTTTTTGTACTTCATTTACACACGAACTCAACGCTGTAAC
Mini-26/26c	GGGAGACAAGAATAAACGCTGAACTATAGTGGATTAACAATACAAGAGACAAG
	ATGCGCCGTATCTTGTGTTGTTAATCCACTATACACGAACTCAACGCTGTAAC
Mini-27/27	GGGAGACAAGAATAAACGCTGAACTATAGTGGATTAAATTTAAATCAGGACAAG
	ATGCGCCGTACTGGTTTAAATTTAATCCACTATACACGAACTCACGCTGTAAC
Mini-for	TAATACGACTCACTATAGGGAGACAAGAATAAACGCT
Mini-rev	GGTACATATTTAGGTTTGTTACAGCGTTGAGTTCGTG

the DNA sequence of *N. meningitidis* strain MC58) were used as primers to amplify the *rnc* coding region from *N. meningitidis* strain B1940. The amplified region extends from residue 711 321 to residue 712 039 in MC58 DNA. Amplimers were cleaved with *Pst*I and *Hin*dIII and inserted downstream from the T7 phage A10 promoter between the *Pst*I and *Hin*dIII sites of the pRSETB vector (Invitrogen), to direct the synthesis of transcripts eventually translated by reticulocyte cell extracts (T&T Kit; Promega).

RNA processing assays

Uniformly ^{32}P-labelled RNAs were obtained by transcribing *in vitro* linear DNA templates with either T7 or Sp6 RNA polymerase as described in [20]. Templates were obtained by PCR amplification of *N. meningitidis* DNA with 50-mers that included either T7 or Sp6 RNA polymerase promoters in their 5′ half, and residues complementary to DNA sequences of interest in their 3′ half. Mini-*nemis* DNA templates were obtained by annealing either 106- or 107-mers containing *nemis* TIRs to the mini-for oligonucleotide (oligomers used are listed in Table 1). Annealed moieties were made double-stranded by the Klenow enzyme, and amplified by PCR by using as primers the oligonucleotides mini-for and mini-rev (see Table 1). Amplimers obtained were checked by DNA sequence analysis. The T7 promoter sequence included in the mini-rev oligonucleotide allowed *in vitro* synthesis of RNA substrates for processing assays. Degradation assays with either bacterial or reticulocyte cell extracts expressing *N. meningitidis* RNase III were carried out as reported in [20]. In processing assays carried out under low salt conditions, both KCl and NH_4Cl were omitted from the reaction buffer.

RNA analyses

Total bacterial RNA was purified on an RNeasy column (Qiagen). Primer extension analyses were performed as reported in [21]. For RNase protection assays, 5 μg of total RNA was mixed with ^{32}P-labelled antisense RNA probes in 30 μl of hybridization buffer (75 % formamide, 20 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl, 0.1 % SDS) and kept at 45 °C for 16 h. After a 1 h incubation at 33 °C with RNase T_1 (2 μg/ml), samples were treated with proteinase K (50 μg/ml) for 15 min at 37 °C, extracted once with phenol, precipitated with ethanol, resuspended in 80 % formamide and loaded on to 6 % (w/v) polyacrylamide/8 M urea gels.

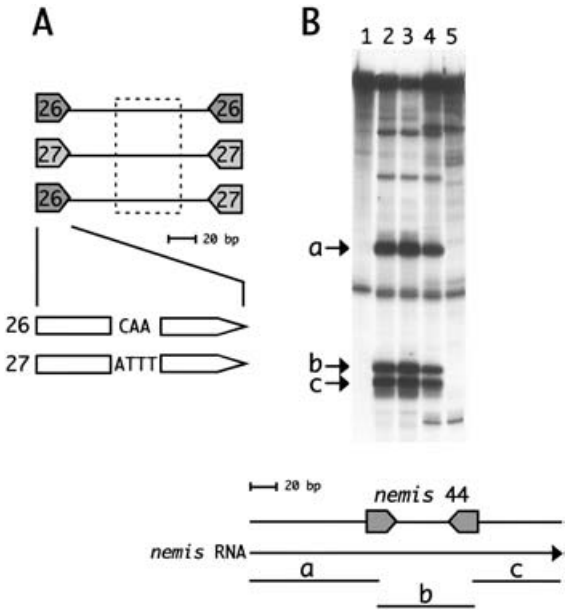


Figure 1 Organization of *nemis* repeats and cleavage of *nemis*$^+$ transcripts

(**A**) *Nemis* feature TIRs that measure, including the TA dinucleotide target duplicated upon genomic insertion, either 26 or 27 bp. TIR types differ in the presence of alternative central DNA stretches. The 50 bp central region found only in long elements is highlighted. (**B**) Approx. 50 000 c.p.m. of radiolabelled RNA spanning *nemis* 44 was incubated for 20 min at 37 °C with approx. 2 μg of two different preparations of crude *N. lactamica* cell extracts (lanes 2 and 3) or *N. meningitidis* RNase III expressed *in vitro* (lane 4). As control, the RNA probe was incubated either alone (lane 1) or with the same amount of reticulocyte extract used to express the recombinant protein (lane 5). Reaction products were separated on a 6 % (w/v) polyacrylamide/8 M urea gel. Bands a–c indicate processed RNA species.

^{32}P-labelled antisense RNA probes were transcribed *in vitro* by Sp6 RNA polymerase.

RNA band-shift experiments were carried out essentially as described in [22]. Aliquots (≈2000 c.p.m.) of gel-purified ^{32}P-labelled RNA (25–50 pmol) were incubated with either whole bacterial extracts or recombinant *N. meningitidis* RNase III in 15 μl reactions, in the presence of 40 mM Hepes (pH 7.2), 20 mM $(NH_4)_2SO_4$, 15 mM potassium acetate, 10 % (v/v) glycerol, 50 mg/ml yeast tRNA and 0.5 mg/ml BSA. Samples were incubated for 30 min at 4 °C and then loaded on to 5 % (w/v) polyacrylamide gels. Electrophoresis was carried out at 4 °C and 23 mA for 2 h in 0.5 × Tris/borate buffer.

RESULTS

Nemis RNA is cleaved by *N. meningitidis* RNase III synthesized *in vitro*

Members of the *nemis* family range in size from 106–108 to 156–158 bp due to the presence/absence of a 50 bp internal segment. Both long and short elements feature TIRs which, including the duplicated TA targets, measure either 26 or 27 bp. Both types of TIR, apart from a short central DNA segment, have the same sequence (Figure 1A). Elements inserted close to coding regions are co-transcribed with flanking ORFs into mRNAs processed at double-helical structures, formed by paired TIRs [13]. Cleavage assays carried out with whole *E. coli* cell extracts support the notion that RNase III is involved in the processing event [20]. The

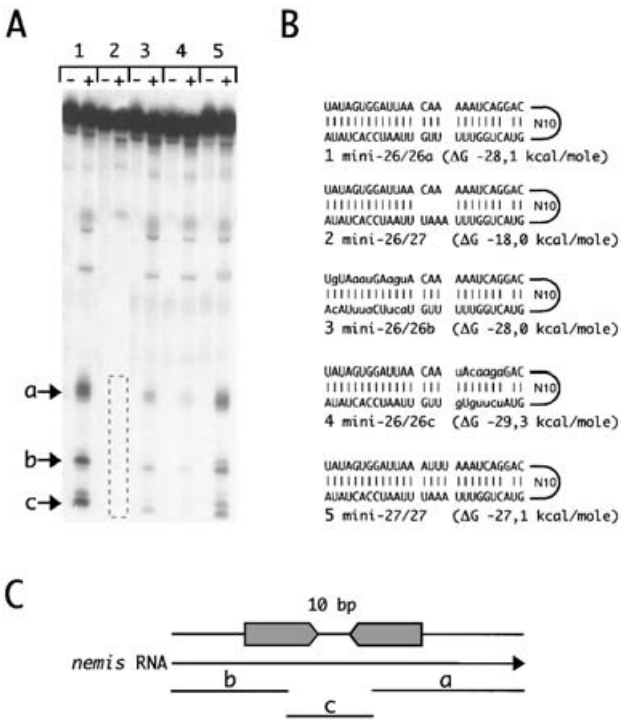


Figure 2 Cleavage of mini-*nemis* RNAs

(**A**) Radiolabelled RNAs spanning mini-*nemis* templates 1–5 carrying different types of TIRs were incubated for 20 min at 37 °C either alone or with *N. meningitidis* RNase III expressed *in vitro* (lanes − and + respectively). Reaction products were separated on a 6 % polyacrylamide/8 M urea gel. The lack of cleavage of the mini-26/27 transcripts is highlighted. (**B**) The degree of complementarity and free energies [23] of the hairpins formed by each transcript are shown (1 kcal = 4.184 kJ). Lower-case letters denote residues substituting for original *nemis* residues. (**C**) RNA species labelled a–c in (**A**) are sketched.

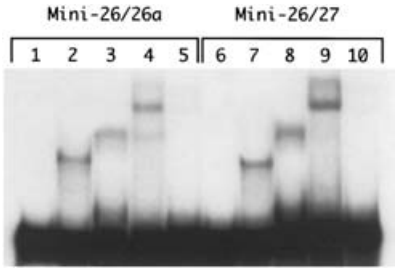


Figure 3 Electrophoretic mobility shift assays of RNA–protein complexes

Radiolabelled RNAs derived from mini-26/26a and mini-26/27 templates (approx. 3000 c.p.m.) were incubated for 20 min at 37 °C either alone (lanes 1 and 6) or with recombinant *N. meningitidis* RNase III (lanes 2 and 7) or with 2 μg of crude *N. lactamica* cell extracts (lanes 3 and 8). The same amount of probe was incubated with 2 μg of crude *E. coli* extracts from *rnc*$^+$ (lanes 4 and 9) and *rnc*$^-$ (lanes 5 and 10) cells. Reaction products were electrophoresed on non-denaturing 5 % (w/v) polyacrylamide gels.

N. meningitidis rnc gene encoding RNase III was amplified by PCR, and used to direct synthesis of the enzyme in reticulocyte cell extracts. As shown in Figure 1(B), the cleavage of *nemis*$^+$ transcripts by Neisserial whole-cell extracts (lanes 2 and 3) is fully mimicked by the recombinant *N. meningitidis* enzyme (lane 4). The low-molecular-mass band visible in lanes 4 and 5 probably results from degradation of the input RNA by nuclease activities found in the reticulocyte extract.

The data reported in Figure 1(B) demonstrate that RNase III is uniquely responsible for the processing of *nemis*$^+$ transcripts, being sufficient to both recognize and cleave *nemis* RNA targets.

The completely sequenced genomes of *N. meningitidis* strains Z2491 and MC58 [4,5] host approx. 250 *nemis* carrying TIRs at both ends. Taking into account the relative distance from neighbouring ORFs, it is plausible that about half of these elements are transcribed along with neisserial genes, and hence are targeted, at the RNA level, by RNase III. To ascertain whether transcripts with the same backbone spanning 26/26, 26/27 and 27/27 *nemis* are processed with comparable efficiency, mini-*nemis* templates were engineered by a PCR-based approach. T7 RNA polymerase-driven transcripts, in which TIR sequences of either the same or different types are separated by 10 nucleotides, were used as substrates for processing experiments (Figure 2). RNA stem–loop structures formed by pairing of either 26 or 27 nt TIRs were cleaved efficiently (Figure 2, lanes 1 and 5 respectively). In contrast, RNA hairpins formed by the pairing of 26 and 27 nt TIRs were fully refractory to RNase III-mediated cleavage (Figure 2, lane 2). The same results (not shown) were obtained by carrying out processing assays at low salt concentrations, a condition known to enhance RNase III-mediated cleavage at secondary sites [24]. In the mini-26/27 transcripts, the double-helical RNA structure formed by TIRs is partly disrupted by an asymmetrical internal loop, and this probably inhibits cleavage. These data were unexpected, since RNA stems of 12 bp are cleaved efficiently by *E. coli* RNase III *in vitro* [24,25], and mini-26/27 transcripts can form a 13 bp stem (see Figure 2). Moreover, substrates exhibiting internal loops are cleaved by the *E. coli* enzyme at one or both sides of the double-helical structure [24,25], and the mini-26/27 transcripts were actually processed by *E. coli* crude extracts (results not shown). Transcripts in which the sequence, but not the strength, of the RNA hairpin was modified were also processed, although to a different extent (Figure 2, lanes 3 and 4). Thus the crucial requirement for cleavage of *nemis*$^+$ transcripts is the degree of complementarity of TIRs.

Interactions between RNase III and *nemis* RNA hairpins

By choosing experimental conditions under which RNase III activity was inhibited, the mini-*nemis* transcripts described in Figure 2 were used as probes in electrophoretic mobility shift assays. As shown in Figure 3, RNAs susceptible and refractory to RNase III-mediated cleavage both formed retarded complexes that had the same mobility upon incubation with either recombinant RNase III (lanes 2 and 7) or neisserial whole-cell extracts (lanes 3 and 8), indicating that the binding of RNase III to *nemis*$^+$ RNA targets may occur without concomitant cleavage. As in RNase III from other micro-organisms, the catalytic and double-stranded RNA-binding domains of the *N. meningitidis* enzyme are located within the N-terminal two-thirds and the C-terminal one-third of the molecule respectively. Functional uncoupling of the two domains has been reported for the *E. coli* enzyme [26]. The sizes of the RNA–protein complexes present in lanes 2 and 7 and lanes 3 and 8 of Figure 3 suggest that RNase III activity in whole-cell extracts is found predominantly in high-molecular-mass complexes. When RNA probes were challenged with *E. coli* whole extracts, retarded electrophoretic complexes were formed with *rnc*$^+$ (Figure 3, lanes 4 and 9) but not with *rnc*$^-$ (Figure 3, lanes 5 and 10) extracts, which lack RNase III because of a Tn10 insertion in the *rnc* gene encoding the enzyme (see [26]). Data indicate that complexes are formed only in the presence of a functional RNase III, and suggest that, unless present specifically in meningococci, cellular factors other than RNase III do not interact directly with *nemis* RNA.

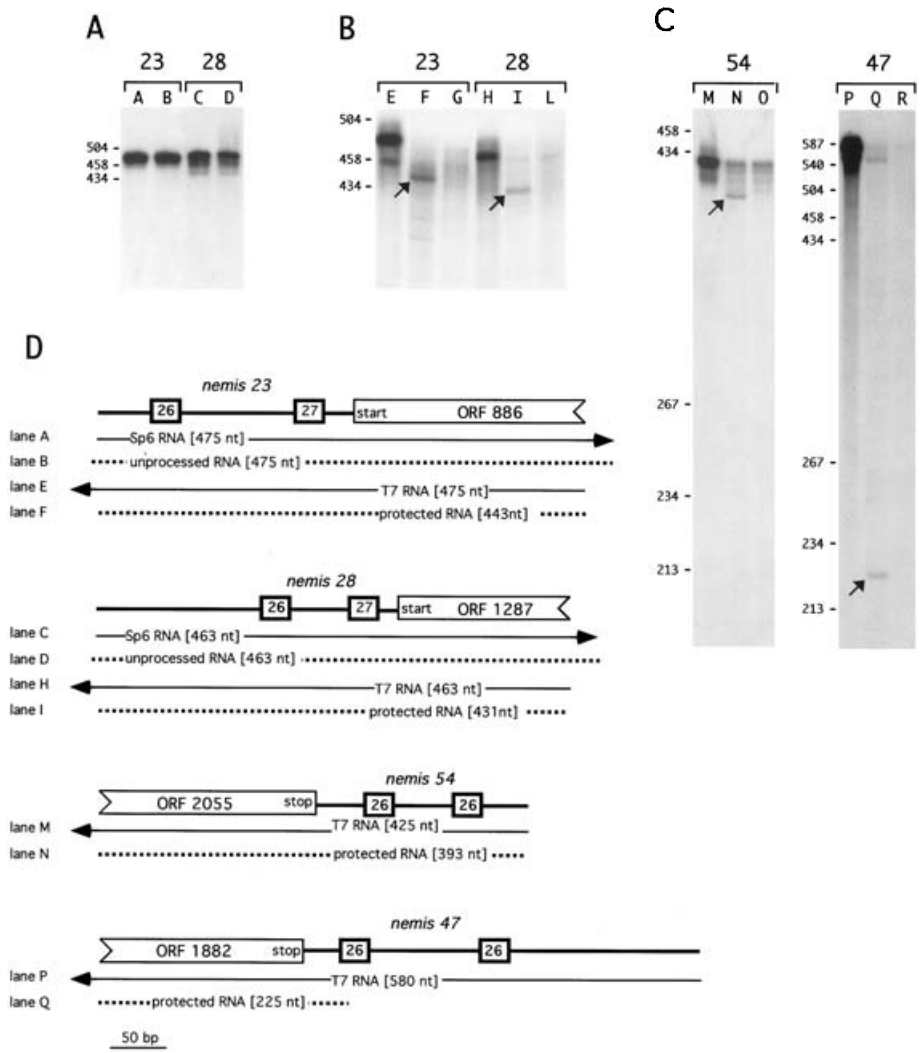


Figure 4 Cellular *nemis*$^+$ transcripts refractory to RNase III-mediated cleavage

(**A**) Uniformly ^{32}P-labelled RNAs transcribed *in vitro* by Sp6 RNA polymerase spanning *nemis* 23 and 28 were incubated either alone (lanes A and C) or with 2 μg of *N. lactamica* cell extracts (lanes B and D). Reaction products were separated on a 6 % polyacrylamide/8 M urea gel. Numbers on the left refer to the size in nucleotides of co-electrophoresed DNA molecular size markers. (**B**) Uniformly ^{32}P-labelled antisense RNAs transcribed *in vitro* by T7 RNA polymerase spanning *nemis* 23 and 28 were hybridized to 5 μg of either total RNA from *N. meningitidis* strain BL859 (lanes F and I) or yeast RNA (lanes G and L). T1 RNase-resistant hybrids were electrophoresed on a 6 % polyacrylamide/8 M urea gel, along with aliquots of untreated input probes (lanes E and H). Major reaction products are marked by arrows. DNA molecular size markers are as in (**A**). (**C**) Uniformly ^{32}P-labelled antisense RNA transcribed *in vitro* by T7 RNA polymerase spanning *nemis* 54 and 47 were hybridized to 5 μg of either total RNA from *N. meningitidis* strain BL859 (lanes N and Q) or yeast RNA (lanes O and R). T1 RNase-resistant hybrids were electrophoresed on a 6 % polyacrylamide/8 M urea gel, along with aliquots of untreated input probes (lanes M and P). (**D**) Template DNAs spanning *nemis* 23, 28, 54 and 47 and part of the neighbouring ORFs are shown. The corresponding sense and antisense transcripts (solid lines; the direction of transcription is indicated by arrows), the reaction products (dotted lines) and their sizes are shown. Transcribed Sp6 and T7 sequences in radiolabelled RNAs are highlighted. ORFs are numbered as in the *N. meningitidis* MC58 strain [5].

Nemis$^+$ mRNAs sensitive and refractory to RNase III-mediated cleavage

To verify results obtained with the mini-transcripts, chromosomal regions encompassing two 26/27 *nemis* (repeats 23 and 28) were amplified by PCR with primers that included phage promoters to allow *in vitro* synthesis of RNAs. The latter were used either as substrates for degradation assays or as antisense probes for RNase protection analyses. We were unable to detect *in vitro* processing of transcripts spanning *nemis* 23 and 28 incubated with *N. lactamica* cell extracts (Figure 4A, lanes B and D). These results contrast sharply with data reported in Figure 1(B), showing that transcripts spanning the 26/26 element 44 are processed *in vitro* by *N. lactamica* cell extracts at double-helical structures formed by paired *nemis* TIRs. Accordingly, *N. meningitidis* mRNAs spanning *nemis* 23 and 28, and ORFs 886 and 1287, located downstream of each, were found to be uncleaved when analysed by RNase T1 mapping (Figure 4B). Thus transcripts spanning 26/27 *nemis* repeats seem also to be refractory to cleavage by RNase III in their natural RNA context.

RNA hairpins formed by 26/26 *nemis* may also be resistant to RNase III-mediated cleavage. Around a dozen *N. meningitidis* mRNAs spanning 26/26 *nemis* monitored by primer extension had been found to be processed at *nemis* TIRs ([13]; see also Figure 5). However, we found that mRNAs spanning ORF 2055, which codes for a transcriptional regulator of the metR family, and *nemis* element 54 were refractory to RNase III-mediated cleavage *in vivo* (Figure 4C, left panel). As a control, RNase protection

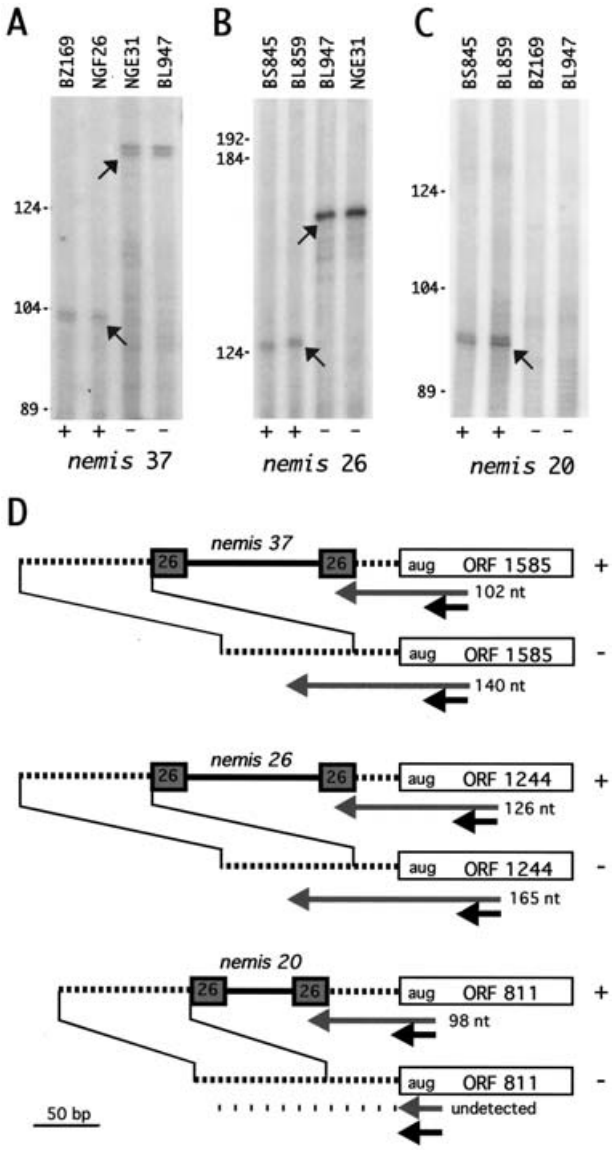


Figure 5 Analysis of corresponding *nemis*$^+$ and *nemis*$^-$ transcripts

(**A**)–(**C**) ^{32}P-5′-end-labelled primers complementary to the coding regions of ORFs 1585 (**A**) 1244 (**B**) and 811 (**C**) were hybridized to total RNA (5 μg) from *N. meningitidis* strains [16] that differ in the presence (+) or absence (−) of *nemis* 37, 26 and 20 upstream of the indicated ORFs. Annealed primer moieties were extended, in the presence of NTPs, by AMV (avian myeloblastosis virus) reverse transcriptase. Reaction products were electrophoresed on a 6 % polyacrylamide/8 M urea gel. Major reaction products are marked by arrows. Numbers to the left of each autoradiogram refer to the size in nucleotides of co-electrophoresed DNA molecular size markers, as in Figure 4. (**D**) The structures of corresponding mRNAs either containing (+) or lacking (−) *nemis* 37, 26 and 20 analysed by primer extension in (**A**)–(**C**) are shown diagrammatically. Dotted lines denote 5′ untranslated transcript regions. Primers and elongation products are denoted by black and grey arrows respectively. The sizes of the reaction products detected in the autoradiograms shown in (**A**)–(**C**) are given.

data obtained by monitoring mRNAs spanning ORF 1882 and the 26/26 *nemis* 47, which were found to be cleaved by RNase III *in vivo*, are shown in the right panel of Figure 4(C). The presence of 26 bp TIRs at the termini of *nemis* element 54 found in strain NGH26, from which RNA was isolated, was confirmed by sequence analyses, ruling out the possibility that our data may reflect changes in the structural organization of *nemis* 54. While the possibility cannot be excluded that a fraction of mRNAs spanning *nemis* 54 are processed by RNase III but cannot be

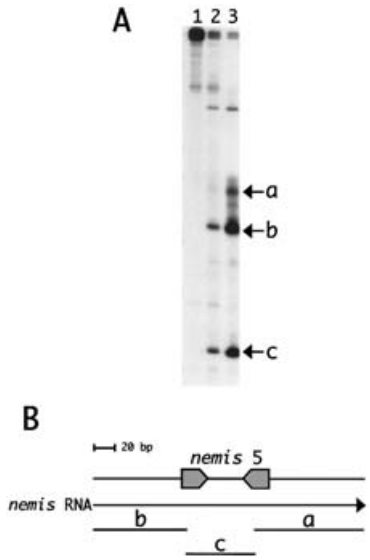


Figure 6 Different stabilities of *nemis* RNA cleavage products

(**A**) Uniformly ^{32}P-labelled RNA spanning *nemis* 5 (lane 1) was incubated either with *N. lactamica* cell extracts (lane 2) or with *N. meningitidis* RNase III expressed *in vitro* (lane 3). Reaction products were separated on a 6 % polyacrylamide/8 M urea gel. (**B**) RNA species labelled a–c in (**A**) are sketched.

detected because they are rapidly degraded, the data suggest that the overall RNA context in which *nemis* RNA hairpins are embedded may result, in some instances, in them being inhibitory for RNase III-mediated cleavage. This also means that the number of unprocessed *nemis*$^+$ mRNAs may be higher than predicted from the structure of RNA hairpins formed by the pairing of *nemis* termini.

Analysis of *nemis*$^+$ transcripts *in vivo*

The identification of *N. meningitidis* strains that lack *nemis* repeats at specific chromosomal locations [16] has enabled us to monitor the steady-state levels of corresponding *nemis*$^+$ and *nemis*$^-$ mRNAs *in vivo*. As shown by primer extension analyses, ORF 1585 transcripts accumulated at comparable levels in both *nemis*$^+$ and *nemis*$^-$ strains (Figure 5A). In contrast, the processed mRNAs encoding ribulose phosphate epimerase (ORF 1244) accumulated approximately 5 times less abundantly than the corresponding unprocessed transcripts (Figure 5B). The levels of *murB* (ORF 811) transcripts were relatively high in the *nemis*$^+$ strains BS845 and BL859, but barely detectable in the *nemis*$^-$ strains BZ169 and BL947 (Figure 5C). Although similar data were obtained by the analysis of independent strains, the possibility cannot be formally ruled out that either the level of synthesis or the overall stability of the various transcripts may vary among the strains analysed. However, quantitative differences emerging from data shown in Figure 5 epitomize the uncertainity in predicting the fate of *nemis*$^+$ mRNAs, since RNase III-mediated cleavage may increase the stability, as well as accelerate the decay, of specific transcripts [27,28].

Differences in the segmental stability of processed *nemis*$^+$ RNA species are clearly illustrated by the *in vitro* data. Transcripts spanning the 26/26 *nemis* 13 were challenged with either *N. meningitidis* RNase III or whole-cell extracts from *N. lactamica* (Figure 6). Cleavage by the recombinant enzyme yielded RNA species that accumulated at comparable levels (see bands a–c in lane 3 of Figure 6). However, when processing was carried out

with whole-cell extracts, band a was barely detectable (Figure 6, lane 2). This contrasts with data reported in Figure 1(B), in which the cleavage products of transcripts spanning the 26/26 *nemis* 44 all exhibited the same relative stability, regardless of whether cleavage was by RNase III synthesized *in vitro* or by neisserial whole-cell extracts.

DISCUSSION

nemis are a major class of repetitive DNA elements that are spread throughout the genomes of pathogenic *Neisseriae*. According to genomic surveys carried out with a representative set of repeats, *nemis* are conserved in meningococci of different sequence types and geographical origins, but are under-represented in strains of the apathogenic *N. lactamica* species [16]. However, genetic exchanges occur at high frequency among clones of the two species, which both colonize the human nasopharynx [29]. Meningococci may have a selective advantage in retaining *nemis* DNA at specific chromosomal sites, and we believe that this correlates with the possible role of several *nemis* in post-transcriptional control.

Analogous to the short *E. coli* intergenic DNA sequences called REPs (repetitive extragenic palindromes) [30–32], *nemis* can be co-transcribed with neighbouring genes. However, RNAs spanning the two classes of repeats enter different degradative pathways: REP sequences are substrates for the mRNA degradative machine [28], while *nemis* sequences are targeted by RNase III. In prokaryotes, this endoribonuclease assists in the maturation of stable RNAs such as the 16 S and 23 S rRNAs. The fate of a few specific mRNAs is also influenced by RNase III-mediated cleavage [27,28].

In order to process *nemis*$^+$ transcripts, RNase III must neither enter into a macromolecular complex nor associate with other cellular components (Figure 1B). Hairpins formed by artificial *nemis*$^+$ transcripts are all cleaved by RNase III, although to different extents (Figure 2). The distance that separates *nemis* TIRs is not relevant for processing. The data also indicate that the degree of complementarity of TIRs, rather than a specific sequence content, is crucial in whether processing occurs. Transcripts spanning elements featuring the same termini (26/26 and 27/27 *nemis*) are cleaved efficiently. In contrast, the bulge in the RNA hairpins formed by TIRs of 26/27 elements results in structures that are resistant to RNase III-mediated cleavage, both *in vitro* and *in vivo* (Figures 2 and 4 respectively). The data reflect an inherent difference between the *N. meningitidis* and the *E. coli* enzymes, since the 26/27 *nemis* RNA substrates can be processed *in vitro* by *E. coli* RNase III. In wholly sequenced neisserial genomes, the 26/27 repeats account for approximately one-quarter of the *nemis* family [13]. *nemis*$^+$ mRNAs refractory to cleavage may be more abundant than those predicted by RNA folding data emerging from *in silico* analyses. Neisserial genomes undergo frequent remodelling, and 'cleavable' 26/26 *nemis* may be replaced by 'uncleavable' 26/27 repeats at any intergenic *nemis*$^+$ site. Examples of such rearrangements are provided by *in silico* comparisons of corresponding regions in the genomes of strains Z2491 and MC58 ([13]; P. P. Di Nocera, unpublished work). Moreover, the overall RNA context in which *nemis* RNA hairpins are embedded may also influence cleavage, and transcripts spanning cleavable *nemis* targets may not be processed *in vivo* (Figure 4C).

Cleavable and uncleavable RNA hairpins formed by *nemis* TIRs both bind RNase III (Figure 3). It has been calculated that fewer than 500 molecules of the enzyme are present in *E. coli* [26]. Should this estimate also apply to *N. meningitidis*, uncleaved *nemis* RNA hairpins may help in tuning RNase III levels by enzyme titration. The hairpin-bound inactive molecules may also provide a 'road block' to 3′ exonuclease activities. Susceptibility or resistance to RNase III-mediated cleavage may either delay or facilitate the turnover of *nemis*$^+$ transcripts. This point is supported by differences both in the steady-state levels of corresponding *nemis*$^+$ and *nemis*$^-$ mRNAs among *N. meningitidis* strains *in vivo* (Figure 5) and in the relative stabilities of the cleavage products of distinct *nemis*$^+$ RNAs *in vitro* (Figures 1B and 6; see also the Experimental section). The set of mRNAs 'decorated' by *nemis* sequences also includes transcripts spanning 26/26 and 27/27 elements which are partly cleaved at *nemis* TIRs, probably due to subtle variations in the structure of RNA stems formed by *nemis* inserted in different orientations [13]. ORFs and elements located either 60–70 bp upstream or 20–30 bp downstream are likely to be co-transcribed. Accordingly, the number of *N. meningitidis* genes potentially regulated by the formation of *nemis* hairpins within their mRNAs could be approx. 100. Of these, many encode transcriptional and/or regulatory factors, and others encode proteins known or hypothesized to be involved in pathogenesis [15,16]. The hypothesis that *nemis*-mediated RNA processing may be relevant in the life cycle of meningococci as pathogenic micro-organisms is substantiated by the knowledge that RNase III, which is not required for viability, is a molecule that is crucial for the survival of meningococci in the infected host [8].

In addition to their role in assembling, re-assorting or disrupting genes, ISs can also activate the expression of neighbouring genes; this generally occurs because of the formation of novel, hybrid promoters at the junction of the IS with the interrupted gene [33]. To our knowledge, *nemis* represent the first example of ISs that are able to influence gene expression by acting as RNA elements. Future analyses should provide a comprehensive list of genes that are co-transcribed along with *nemis* in *N. meningitidis* populations, so clarifying whether the presence/absence of *nemis* repeats at specific chromosomal sites may account for the selective properties of particular pathogenic *N. meningitidis* strains.

We thank Dr Luca Cardone for advice on expressing RNase III *in vitro*. We are indebted to Dr Nina Dathan for critical reading of the manuscript. This work is partly supported by a grant from MURST (Ministero della Università e della Ricerca Scientifica e Tecnologica), PRIN 2002 program, to P. P. D. N.

REFERENCES

1 Meyer, T. F., Pohlner, J. and van Putten, J. P. (1994) Biology of the pathogenic Neisseriae. Curr. Top. Microbiol. Immunol. **192**, 283–317
2 Achtman, M. (1995) Global epidemiology of meningococcal disease, In Meningococcal Disease (Cartwright, K., ed.), pp. 159–175, John Wiley & Sons, Chichester
3 Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A. et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. **95**, 3140–3145
4 Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T. et al. (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature (London) **404**, 502–506
5 Tettelin, H., Saunders, N. J., Heidelberg, J., Jeffries, A. C., Nelson, K. E., Eisen, J. A., Ketchum, K. A., Hood, D. W., Peden, J. F., Dodson, R. J. et al. (2000) Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287**, 1809–1815
6 Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R. and Moxon, E. R. (2000) Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. Mol. Microbiol. **37**, 207–215
7 Perrin, A., Bonacorsi, S., Carbonnelle, E., Talibi, D., Dessen, P., Nassif, X. and Tinsley, C. R. (2002) Comparative genomics identifies the genetic islands that distinguish *Neisseria meningitidis*, the agent of cerebrospinal meningitis, from other *Neisseria* species. Infect. Immun. **70**, 7063–7072
8 Sun, Y. H., Bakshi, S., Chalmers, R. and Tang, C. M. (2000) Functional genomics of *Neisseria meningitidis* pathogenesis. Nat. Med. **6**, 1269–1273

9 Maiden, M. C. J., Malorny, B. and Achtman, M. (1996) A global gene pool in the neisseriae. Mol. Microbiol. **21**, 1297–1298
10 Morelli, G., Malorny, B., Müller, K., Seiler, A., Wang, J., del Valle, J. and Achtman, M. (1997) Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. Mol. Microbiol. **25**, 1047–1064
11 Feil, E. J., Maiden, M. C., Achtman, M. and Spratt, B. G. (1999) The relative contribution of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. **16**, 1496–1502
12 Correia, F. F., Inouye, S. and Inouye, M. (1988) A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae.* J. Biol. Chem. **263**, 12194–12198
13 Mazzone, M., De Gregorio, E., Lavitola, A., Pagliarulo, C., Alifano, P. and Di Nocera, P. P. (2001) Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic Neisseriae. Gene **278**, 211–222
14 Buisine, N., Tang, C. M. and Chalmers, R. (2002) Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic Neisseria species. FEBS Lett. **522**, 52–58
15 Liu, S. V., Saunders, N. J., Jeffries, A. and Rest, R. F. (2002) Genome analysis and strain comparison of correia repeats and correia repeat-enclosed elements in pathogenic Neisseria. J. Bacteriol. **184**, 6163–6173
16 De Gregorio, E., Abrescia, C., Carlomagno, M. S. and Di Nocera, P. P. (2003) Asymmetric distribution of *nemis* DNA repeats among pathogenic and apathogenic *Neisseriae*. Infect. Immun. **71**, 4217–4221
17 Black, C. G., Fyfe, J. A. and Davies, J. K. (1995) A promoter associated with the neisserial repeat can be used to transcribe the uvrB gene from *Neisseria gonorrhoeae*. J. Bacteriol. **177**, 1952–1958
18 Francis, F., Ramirez-Arcos, S., Salimnia, H., Victor, C. and Dillon, J. R. (2000) Organization and transcription of the division cell wall (dcw) cluster in *Neisseria gonorrhoeae*. Gene **251**, 141–151
19 Hill, S. A., Samuels, D. S., Carlson, J. H., Wilson, J., Hogan, D., Lubke, L. and Belland, R. J. (1997) Integration host factor is a transcriptional cofactor of pilE in *Neisseria gonorrhoeae*. Mol. Microbiol. **23**, 649–656
20 De Gregorio, E., Abrescia, C., Carlomagno, M. S. and Di Nocera, P. P. (2002) The abundant class of *nemis* repeats provides RNA substrates for ribonuclease III in Neisseriae. Biochim. Biophys. Acta **1576**, 39–44

Received 8 April 2003/16 June 2003; accepted 25 June 2003
Published as BJ Immediate Publication 26 June 2003, DOI 10.1042/BJ20030533

21 Abrescia, C., De Gregorio, E., Frontini, M., Mantovani, R. and Di Nocera, P. P. (2002) A novel intragenic sequence enhances initiator-dependent transcription in human embryonic kidney 293 cells. J. Biol. Chem. **277**, 19594–19599
22 Carlomagno, M. S. and Nappo, A. (2001) The antiterminator NusB enhances termination at a sub-optimal Rho site. J. Mol. Biol. **309**, 19–28
23 Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. Science **244**, 48–52
24 Li, H. L., Chelladurai, B. S., Zhang, K. and Nicholson, A. W. (1993) Ribonuclease III cleavage of a bacteriophage T7 processing signal. Divalent cation specificity, and specific anion effects. Nucleic Acids Res. **21**, 1919–1925
25 Nicholson, A. W. (1999) Function, mechanism and regulation of bacterial ribonucleases. FEMS Microbiol. Rev. **23**, 371–390
26 Dasgupta, S., Fernandez, L., Kameyama, L., Inada, T., Nakamura, Y., Pappas, A. and Court, D. L. (1998) Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNase III – the effect of dsRNA binding on gene expression. Mol. Microbiol. **28**, 629–640
27 Court, D. L. (1993) RNA processing and degradation by RNase III. In Control of Messenger RNA Stability (Belasco, G. J. and Brawerman, G., eds.), pp. 71–116, Academic Press, San Diego
28 Coburn, G. A. and Mackie, G. A. (1999) Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. Prog. Nucleic Acid Res. Mol. Biol. **62**, 55–108
29 Linz, B., Schenker, M., Zhu, P. and Achtman, M. (2000) Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*. Mol. Microbiol. **36**, 1049–1058
30 Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D. and Higgins, C. F. (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48**, 297–310
31 Newbury, S. F., Smith, N. H. and Higgins, C. F. (1987) Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell **51**, 1131–1143
32 McLaren, R. S., Newbury, S. F., Dance, G. S., Causton, H. C. and Higgins, C. F. (1991) mRNA degradation by processive 3′-5′ exoribonucleases *in vitro* and the implications for prokaryotic mRNA decay *in vivo*. J. Mol. Biol. **221**, 81–95
33 Mahillon, J., Leonard, C. and Chandler, M. (1999) IS elements as constituents of bacterial genomes. Res. Microbiol. **150**, 675–687