

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

03062
2ej. 7
RECIBIDA EN
BIBLIOTECA
MAY 30 1988
SALA

**ESTUDIO DE LAS SECUENCIAS PALINDROMICAS REPETIDAS
EXTRAGENICAS (REP) EN EL GENOMA PROCARIONTE.**

Tesis que presenta Enrique Merino Pérez
para obtener el grado de Maestro en
Investigación Biomédica Básica

Cuernavaca, Mor. 1988.

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

TESIS CON FALLA DE ORIGEN

I.-	INTRODUCCION	
1.1	Ingeniería Genética	1
1.2	Características de la secuencia REP	2
II.-	ANTECEDENTES Y OBJETIVOS	
	Antecedentes	
2.1	Posibles funciones de la secuencia REP	5
2.1.1	Regulación de la expresión génica	5
2.1.2	Rearreglos cromosomales y diversidad génica	8
2.1.3	Estructura y organización del cromosoma	8
2.1.4	DNA egoísta en el genoma bacteriano	9
2.2	Estabilidad del mRNA bacteriano	9
2.3	Gene estructural de la glutamato deshidrogenasa de <u>Escherichia coli</u> K-12	12
	Objetivos	12
III.-	MATERIALES Y METODOS	14
IV.-	RESULTADOS	
4.1	Actividad específica de la enzima GDH y estabilidad del mRNA de <u>gdhA</u>	20
4.2	Búsqueda por computadora de secuencias REP	21
4.3	Búsqueda por computadora de otras secuencias reiteradas	23
V.-	DISCUSION	32
VI.-	BIBLIOGRAFIA	41
VII.-	Apendice I. PROGRAMAS DE COMPUTADORA DESARROLLADOS	
7.1	Programa 'Localiza.cadenas'	48
7.2	Programa 'Localiza.REPs'	54
7.3	Programa 'Elimina.redundancia'	59
7.4	Programa 'Busqueda.Hash'	61
VII.-	Apendice II. PUBLICACIONES	65

I.-	INTRODUCCION	
1.1	Ingenieria Genética	1
1.2	Características de la secuencia REP	2
II.-	ANTECEDENTES Y OBJETIVOS	
	Antecedentes	
2.1	Posibles funciones de la secuencia REP	5
2.1.1	Regulación de la expresión génica	5
2.1.2	Rearreglos cromosomales y diversidad génica	8
2.1.3	Estructura y organización del cromosoma	8
2.1.4	DNA egoísta en el genoma bacteriano	9
2.2	Estabilidad del mRNA bacteriano	9
2.3	Gene estructural de la glutamato deshidrogenasa de <u>Escherichia coli</u> K-12	12
	Objetivos	12
III.-	MATERIALES Y METODOS	14
IV.-	RESULTADOS	
4.1	Actividad específica de la enzima GDH y estabilidad del mRNA de <u>gdhA</u>	20
4.2	Búsqueda por computadora de secuencias REP	21
4.3	Búsqueda por computadora de otras secuencias reiteradas	23
V.-	DISCUSION	32
VI.-	BIBLIOGRAFIA	41
VII.-	Apendice I. PROGRAMAS DE COMPUTADORA DESARROLLADOS	
7.1	Programa 'Localiza.cadenas'	48
7.2	Programa 'Localiza.REPs'	54
7.3	Programa 'Elimina.redundancia'	59
7.4	Programa 'Busqueda.Hash'	61
VII.-	Apendice II. PUBLICACIONES	65

En la actualidad, gracias a esta metodología se ha podido determinar la secuencia nucleotídica del genoma de diferentes virus y de un gran porcentaje de algunos microorganismos. El análisis de dichas secuencias mediante programas de computadora, así como experimentos de reasociación de moléculas de DNA, han revelado que los organismos en general poseen secuencias únicas constituyendo una parte importante del genoma (10, 11, 14). Sin embargo, los organismos eucariotes poseen un porcentaje elevado de su genoma como regiones de DNA que se repiten muchas veces. Por otro lado, dicha reiteración del material genético, no ha sido observada en organismos procariotes, cuya tendencia general en este aspecto es la de poseer un genoma de menor tamaño carente de secuencias altamente repetidas (14).

1.2 Características de la secuencia REP

A pesar de la tendencia general del genoma procarionte de no poseer secuencias altamente reiteradas, estudios recientes de los genomas de Escherichia coli y Salmonella typhimurium (34, 74, 25, 4) han revelado la existencia de una secuencia palindrómica excepcionalmente reiterada. Esta secuencia se le ha denominado unidad palindrómica PU (25) o secuencia REP (Repetitive Extragenic Palindromic sequence) (74, 4), atendiendo al hecho de haber sido localizada exclusivamente en regiones extragénicas. Las características más importantes de estas secuencias, que son descritas en los artículos que se anexan a la presente tesis, se mencionan continuación:

- Su longitud aproximada es de 35 pares de bases.

- Aún y cuando se han realizado diversas búsquedas con programas de computadora en bancos de datos de ácidos nucleicos, sólo ha sido posible encontrar secuencias REP en los genomas de Escherichia coli y Salmonella typhimurium (74, 25, 27, 28).

- Se encuentran altamente conservadas ya que difieren unas de otras en tan sólo unos cuantos pares de bases. (ver tablas II y III de Becerril et al., 1985).

- Se encuentran altamente reiteradas; se calcula que existan aproximadamente 500 copias de esta secuencia por genoma (74)

- Están presentes en diferentes tipos de operones entre los cuales se encuentran catabólicos (lacY), biosintéticos (hisG), de transporte (hisJ), respiratorios (ndhA), y regulatorios (trpR) (74).

- La orientación, posición y distribución de estas secuencias, parece no obedecer a ninguna regla conocida.

- En todos los genes que poseen secuencias REP en donde el término de la transcripción es conocido, éstas se encuentran formando parte del transcrito.

- Debido a su propiedad de palíndromía, las secuencias REP poseen una alta capacidad para formar estructuras secundarias de tallo-asa (ver Fig 4, Becerril et al., 1985). Al parecer la palíndromía es una propiedad importante en estas secuencias. Así lo indica el hecho de que a pesar de la aparente flexibilidad en seis de las posiciones del consenso de esta secuencia, la capacidad del apareamiento de bases es mantenida (ver tablas II y III de Becerril et al., 1985 y Fig. 1 de Merino y Bolívar, 1987c). Aún más, puede también observarse que cuando una de las

bases del primer brazo del palíndromo difiere del consenso, la base correspondiente en el segundo brazo también difiere, de tal forma que su apareamiento sea posible (74).

- Las secuencias REP existen en mono o multicopias adyacentes, en cuyo caso, estructuras secundarias alternativas pueden ser formadas en los transcritos que las porten. (ver Fig. 1 y Fig 4 de Becerril et al., 1985).

- Todas las secuencias REP hasta ahora reportadas son extragénicas (con excepción de una secuencia REP ubicada en el gene estructural de la enzima ribonucleósido difosfato reductasa (nrDA) de Escherichia coli, encontrada en el presente estudio. Ver Merino y Bolívar, 1987b y el capítulo V de esta tesis).

- En la mayoría de los casos, las secuencias REP están ubicadas en la región 3' terminal del gene.

- Cuando las secuencias REP se encuentran en regiones intercistónicas de transcritos policistónicos, la actividad específica del producto peptídico de los genes que se encuentran antes de esta secuencia, es mayor respecto al producto génico de los cistrones ubicados después de la misma (74, 27, 28).

- Delecciones de estas secuencias del extremo terminal 3' de un gene, originan un decremento en su actividad específica de aproximadamente un 50% (74, 4).

II ANTECEDENTES Y OBJETIVOS

ANTECEDENTES:

2.1 Posibles funciones de la secuencia REP

Actualmente no se conoce con certeza la posible función o funciones de la secuencia REP en el genoma procarionte. No obstante, el alto grado de similitud entre ellas, así como el hecho de encontrarlas en tan alto número de copias, ha permitido pensar que este elemento pueda tener una función(es) celular(es) importante(s). Con el objeto de entender dicha(s) función(es), se han formulado diversas hipótesis. A continuación se mencionan algunas de ellas:

2.1.1 Regulación de la expresión génica

Se ha propuesto que las secuencias REP pueden ser un nuevo elemento de regulación considerando que alteran la concentración del producto final de algunos genes que se encuentran en su vecindad (74, 4). Se ha pensado que la capacidad de la secuencia REP para formar estructuras secundarias estables, le permita alterar la estabilidad de los transcritos en los que se encuentra, al formar un bloque que impediría parcialmente la degradación (en sentido 3'--->5') por algunas exonucleasas. Este planteamiento podría explicar el porque los elementos REP presentes en mRNA monocistrónicos se encuentran generalmente en el extremo terminal 3'. Los resultados de mapeo de protección del transcrito policistrónico del operón de transporte de histidina en Salmonella typhimurium, contra la acción de degradación de

nucleasa Si, apoyan la anterior hipótesis, ya que aproximadamente un 30% de estos transcritos terminan en el extremo 3' de su segunda secuencia REP (74).

Stern et al., (1984) han propuesto que las secuencias REP pudieran tener un papel opuesto al de estabilizar al mRNA. Basados en el hecho de no haber encontrado secuencias REP en transcritos estables, han sugerido que REP puede disminuir la estabilidad de los transcritos de los que forma parte, al ser sustrato de endonucleasas específicas (i.e. RNasaIII). Sin embargo es conveniente aclarar que la búsqueda por computadora realizada en nuestro laboratorio en bancos de datos de secuencias nucleotídicas, ha permitido encontrar secuencias REP en unidades transcripcionales estables, como lo son las del tyr-tRNA y supBE. Los resultados de dicha búsqueda también han revelado que existen secuencias similares a REP (tipo-REP) formando parte de genes que codifican para las moléculas estables ile-tRNA de Bacillus subtilis y Escherichia coli, así como los rRNA 16S y 23S de Escherichia coli (Fig. 2 de Becerril et al., 1985).

Higgins et al. (1982), han observado que cuando REP se encuentra en unidades transcripcionales policistrónicas, la concentración del producto de los genes que se traducen después de esta secuencia, es menor con relación a la de los que lo hacen antes de la misma. Lo anterior ha permitido pensar que REP pudiera originar tal diferencia al servir como un terminador parcial de la transcripción. Sin embargo, experimentos in vitro

realizados con el operón de transporte de histidina en Salmonella typhimurium, indican que no existe pausa o término de la transcripción en las secuencias REP (74).

Nuestro grupo ha descrito que los 120 primeros nucleótidos a partir del codón de terminación del mRNA del gene gdhA, que incluyen a dos secuencias REP, pueden formar una estructura secundaria parecida a la del glu-tRNA (Becerril et al., 1985). Esta similitud permite suponer que dicha estructura pudiera ser reconocida por un tipo de proteína modificadora, tales como las que actúan sobre los tRNA, que alterara la tasa de traducción de sus correspondientes mRNAs. Existen datos que indican que la posible regulación del producto final de genes cuyos transcritos tienen la capacidad de adoptar estructuras similares a las de tRNA, no es exclusiva del gene gdhA. Ames et al. (1), han reportado que la región líder del operón de histidina en Salmonella typhimurium puede adoptar una estructura similar a la de su correspondiente tRNA, y han sugerido que dicha estructura pudiera estar involucrada en la modulación de la transcripción de este operón, favoreciendo o impidiendo la estructura de atenuación del líder. Observaciones similares han sido reportados por Rufus y Julius (69) en el operón ilvGMEDA de Escherichia coli, Salmonella typhimurium, y Klebsiella aerogenes, en donde la región líder del operón posee una similitud con la secuencia del Leu-tRNA. Paralelamente, estudios realizados por Haenni et al. (32), indican que el RNA de algunos virus, poseen secuencias parecidas a las de los tRNAs que pudieran ser reconocidas y aminoacetiladas por enzimas de su hospedero.

El hecho de que los rRNAs estén relacionados con la unión de proteínas ribosomales (62) y que las secuencias REP o tipo-REP se encuentren presentes, no sólo en moléculas de mRNA, si no también en moléculas de rRNA y tRNA, permite pensar que las secuencias REP pudieran interactuar con proteínas involucradas en los eventos de traducción (4).

2.1.2 Rearreglos cromosomales y diversidad génica

Se ha pensado que, dada la gran similitud que poseen las secuencias REP entre sí, éstas pudieran ser sitios de recombinación de rearrreglos cromosomales, aumentando con é ello la posibilidad de generar diversidad en el genoma (34, 74, 25). La única evidencia en favor de esta hipótesis, es que el término del rearrreglo cromosomal que incluye al operón biosintético de histidina en Salmonella typhimurium, está ubicado en la región intercistronica hisG-hisD, en donde se encuentra una secuencia REP (74).

2.1.3 Estructura y organización del cromosoma

Una posible alternativa de la función de las secuencias REP, es la de intervenir en la estructuración y organización del genoma procarionte. Stern et al. (1984), han sugerido que las secuencias REP pudieran servir como sitio de unión de proteínas específicas que permitan anclar al cromosoma, o bien, delinearlos en dominios funcionales. Basados en que las fuerzas que permiten al cromosoma bacteriano poseer una estructura compacta son fundamentalmente interacciones DNA-RNA (82), estos autores han

propuesto que la forma transcrita de REP pudiera contribuir a tal interacción, hipótesis que es congruente con el dato experimental de que siempre se ha encontrado a REP formando parte de unidades transcripcionales. Recientemente, experimentos de protección en contra de la acción de degradación de la exonucleasa III, han proporcionado las primeras evidencias en favor de esta hipótesis, al demostrar que las secuencias REP son sitio de unión de algunas proteínas asociadas al cromosoma (26).

2.1.4. DNA egoista en el genoma bacteriano

Finalmente, durante el desarrollo de esta tesis, Newbury et al. (59), han propuesto que las secuencias REP no poseen ningún papel celular importante y que se han dispersado en el cromosoma mediante un fenómeno de conversión génica mediado por mRNA. La anterior hipótesis está basada en dos diferentes tipos de evidencias. La primera se refiere al hecho de no haber sido posible encontrar claramente alguna función a las secuencias REP. La segunda evidencia esta basada en que la secuencia REP de algunos operones de Escherichia coli, están ausentes en los operones homólogos de Salmonella typhimurium, y viceversa.

2.2. Estabilidad del mRNA bacteriano

Con el objeto de entender el posible papel de las secuencias REP en la estabilidad del mRNA (ver la sección 2.1.1 del capítulo II), a continuación se mencionan algunos de los aspectos relacionados con dicho fenómeno.

Las propiedades básicas de la cinética de degradación de mRNA en bacterias fueron originalmente descritas por Kepes y colaboradores. Existen dos características que pueden observarse en este proceso de degradación: La primera se refiere a la pérdida de masa del mRNA que puede ser medida por el decremento del número de transcritos durante un intervalo determinado de tiempo. La segunda característica es la pérdida de función o pérdida de capacidad del mRNA para dirigir la biosíntesis de proteínas. Tanto la degradación de masa como la pérdida de función del mRNA, se caracterizan por seguir cinéticas de tipo exponencial (44, 46, 45, 35).

Se ha observado que la vida media de los transcritos de organismos eucariotes es relativamente larga respecto al tiempo de su división celular, a diferencia de las vidas medias de los transcritos de los organismos procariotes, que suelen ser significativamente menores (revisado en 49). Una consecuencia de la rápida tasa de degradación del mRNA en organismos procariotes es que una gran porción de sus transcritos se encuentra en proceso de decaimiento. Así por ejemplo, Kennell y Riedman han reportado que en cualquier momento, cerca de la mitad de los transcritos de beta-Galactosidasa en Escherichia coli se encuentran en proceso de degradación (42).

En Escherichia coli, cada transcrito decae con una tasa específica que puede variar desde 30 seg hasta 8 min a 37°C (8), 40 seg a 20 min a 30°C (61) y 6.5 min a 40 min a 15°C (41). En la actualidad no se cuenta con ningún modelo que permita entender con claridad el fenómeno de degradación del mRNA que pueda

explicar las causas de tales diferencias. Algunos de los aspectos relevantes a este fenómeno que se han podido determinar, son los siguientes:

- La tasa de degradación de un transcrito es independiente de su longitud (8).

- La tasa de degradación de un mensajero no está relacionada con su posición en un transcrito policistrónico (revisado en 49).

- En general, la vida media ^{de} los transcritos de fagos es superior a la del mRNA de la bacteria hospedera, sin embargo es frecuente que dichos transcritos puedan ser inactivados antes que el mRNA sea totalmente degradado (revisado en 49).

- En general, mutaciones sin sentido en la región 5' terminal de un un gene, originan un aumento en la tasa de degradación de sus transcritos correspondientes. Antibióticos que bloquean la traducción tales como el cloramfenicol o el ácido fusídico, comúnmente tienden a estabilizar al mRNA. Lo anteriormente expuesto sugiere que la región 5' del mRNA asociada a ribosomas, puede ser protegida de la degradación. De acuerdo a esta hipótesis, los transcritos con altas tasas de inicio de traducción son mas estables (revisado en 49).

Actualmente no se ha determinado con exactitud las rutas de degradación del mRNA. Estas ^{podrían} pueden ser:

- De un extremo del mRNA al otro, mediante la degradación exonucleolítica en sentido 3'-->5' o 5'-->3'.

- Degradación endonucleolítica en cualquier sitio o en sitios específicos del transcrito.

- Cualquier combinación de los puntos anteriores.

3.3 Gené estructural de la glutamato deshidrogenasa de

Escherichia coli K-12

Trabajos realizados en el proyecto de investigación "Genes del Metabolismo Nitrogenado en Escherichia coli K-12" que se lleva a cabo en el laboratorio del Dr. Francisco Bolivar Zapata en el Centro de Investigación sobre Ingeniería Genética y Biotecnología, UNAM, han detectado la presencia de dos secuencias REP en el extremo terminal 3' del gene estructural de la glutamato deshidrogenasa (gdhA). Se ha determinado que deleciones de una de estas secuencias origina una disminución de la actividad específica de la enzima GDH en aproximadamente un 50% (74). Al momento de iniciar el presente trabajo (octubre de 1985), no se conocía con exactitud la etapa en que se realizaba esta posible regulación.

OBJETIVOS

El presente proyecto de maestría tuvo como objetivo estudiar la(s) posible(s) función(es) de las secuencias REP en el cromosoma bacteriano, así como el posible papel que tienen en la regulación diferencial de los genes en los que se encuentran, tomando como modelo de trabajo al gene estructural de glutamato deshidrogenasa (gdhA) de Escherichia coli K-12 y sus regiones aledañas, cuyas características más importantes son las siguientes:

- Posee dos secuencias REP en direcciones opuestas en su extremo terminal 3'. (4).

- La enzima glutamato deshidrogenasa (GDH) ocupa un papel central en el metabolismo celular, ya que permite la interconversión de alfa-cetoglutarato (que genera equivalentes reductores en el ciclo de Krebs) y glutamato (que es donador primario de grupos amino para la síntesis de aminoácidos por transaminación). Esta interconversión es un punto de cruce en el metabolismo nitrogenado y del carbono.

- El gene estructural y regiones que lo flanquean, están clonados en un vehículo molecular (19).

- Se ha determinado la secuencia completa del gene y regiones que lo flanquean, así como también las posibles regiones involucradas en la regulación de la expresión del gene. (80, 78).

- La actividad de la enzima GDH puede ser medida con facilidad mediante la determinación de la tasa de oxidación de NADPH+, en una mezcla de reacción que contiene alfa-cetoglutarato y NH4+ (19).

- La actividad específica del producto del gene gdhA disminuye en aproximadamente un 50% cuando la segunda de sus secuencias REP es eliminada su extremo terminal 3' (4).

III. MATERIALES Y METODOS:

La mayor parte de los materiales y métodos empleados en el presente estudio, son descritos en las publicaciones que se anexan en el apéndice II. Estas comprenden:

MATERIALES:

- Cepas y Plásmidos:

Descritos en la sección "MATERIALS AND METHODS" inciso "Strains and plasmids" del artículo de Becerril, et al., 1985 y Sección "EXPERIMENTAL AND DISCUSSION" inciso "Construction and mRNA characterization", de Merino et al., 1987a.

- Medios de cultivo:

Descritos en la Fig. 3 de Becerril, et al., 1985 y Fig. 2 de Merino, et al., 1987a.

- Software:

a) Bancos de datos de Ácidos nucleicos:

GenBank, versiones septiembre de 1984 y agosto de 1986.

b) Programas de computadora:

Con el objeto de realizar el estudio teórico de las secuencias REP así como el de otras secuencias altamente repetidas en el genoma bacteriano, fueron desarrollados los programas de computadora "Localiza.REPs", "Elimina.redundancia", "Búsqueda.Hash" y "Localiza.cadenas" cuyos listados se anexan en el apéndice I.

- Hardware

Computadora Burroughs B7800.

Computadora Vax 11/780.

Microcomputadora IBM Pc.

METODOS:

- Transformación:

Las cepas de Escherichia coli utilizadas en el estudio, fueron transformadas según el método de Cohen et al. (17), que básicamente consiste en la permeabilización de la membrana por el tratamiento con CaCl 30mM, seguido de un choque térmico a 42°C.

- Determinación de la actividad de la enzima GDH:

La actividad de la enzima GDH se cuantificó a partir de la tasa de oxidación de NADPH según la técnica reportada por Covarrubias et al. (19).

- Determinación de la cantidad de proteína:

La concentración de proteína en el medio fue determinada utilizando el colorante azul brillante de Coomassie G-250, de acuerdo al método descrito por Bradford (13).

- Purificación de DNA de plásmido:

El método utilizado para la purificación de plásmido de volúmenes grandes (comunmente mayores a 250 ml), fue el de la lisis alcalina descrito por Maniatis et al. (53). La

amplificación del número de copias del plásmido se realizó en la fase logarítmica del cultivo mediante la adición del antibiótico bacterioestático cloranfenicol (16).

- Microensayo de DNA de plásmido:

La obtención de DNA de plásmido de volúmenes pequeños de cultivo (aproximadamente 3 ml), fué realizada por el método de la lisis alcalina descrito por Rodriguez y Tait (68).

- Quantificación del número de plásmidos:

Con el objeto de ajustar proporcionalmente los valores de las actividades de la enzima GDH al número de copias de los plásmidos pSAE4 y pBB5 en las cepas transformadas, se tomaron alicuotas de dichas cepas que contuvieran el mismo número de células. Posteriormente, las alicuotas fueron mezcladas, y de ellas se obtuvo simultáneamente el DNA de ambos plásmidos. Dependiendo de su tamaño, los plásmidos fueron separados electroforéticamente en un gel de agarosa al 1%. El barrido densitométrico del negativo ~~de~~ de la fotografía del gel permitió calcular la cantidad relativa del número de copias de un plásmido respecto al otro.

- Purificación del RNA total:

Para extraer el RNA total de las cepas transformadas con los plásmidos pSAE4 y pBB5 se utilizó el método propuesto por Young y Furano (83), que consiste esencialmente en la degradación específica de DNA mediante la utilización de DNasa libre de RNAsa.

- Cuantificación de RNA:

La cantidad de RNA obtenida en la extracción fue determinada a partir de las lecturas a 260 y 280 nm de acuerdo a lo reportado por Maniatis et al. (53). La integridad del RNA fue verificada en geles desnaturizantes de acrilamida 5% urea 8 M, teñidos con Bromuro de Etidio y observados en un transiluminador de luz ultravioleta (65).

- Electroforesis de RNA en geles de agarosa/formaldehido:

Con el objeto de separar el RNA total en base a su tamaño, las muestras de RNA fueron desnaturizadas en una solución 2.2 M de formaldehido, 50% de formamida, 0.018 M Na_2HPO_4 , 0.002 M NaH_2PO_4 a una temperatura de 60 °C durante 5 min. Posteriormente, colocadas en un gel de 1% agarosa - 2.2 M formaldehido y sometidas a un campo electroforético de 30 V durante 10 hrs de acuerdo a lo reportado por Lehrach et al. (48).

- Marcado radioactivo de la sonda de DNA:

El marcado radioactivo de la sonda de DNA (fragmento del plásmido PBB5 que incluye al gene completo de gdhA) empleada en los experimentos de hibridización, fué llevado a cabo según el método de "Nick-translation" (Traslado de corte), descrito por Rigby et al. (64). Este método hace uso de las propiedades endonucleolíticas de la enzima DNasa I para producir un mellado en una de las cadenas de la sonda, alrededor del cual la enzima DNA polimerasa I, removerá nucleótidos que posteriormente serán reemplazados a partir de nucleósidos, cuyos fosfatos alfa son radioactivos. El isótopo empleado en el marcado fué alfa-³²P-CTP.

- Hibridización de ácidos nucleicos:

Después del fraccionamiento electroforético en gel desnaturante, el RNA es transferido a una membrana (en la referencia 75 se presenta una revisión del uso de diferentes tipos de ellas) de acuerdo a lo reportado por Southern (73). La prehibridización, hibridización y lavados fueron llevados a cabo de acuerdo a lo descrito por Maniatis et al. (53) y se describen en el pie de Fig. 2 del artículo de Merino et al., 1987a (ver apéndice I).

- Determinación de la estabilidad del mRNA de gdhA:

La determinación de la vida media del mRNA incluye en forma conjunta, casi todas las metodologías anteriormente señaladas. Particularmente la estabilidad de los mRNA gdhA transcritos de los plásmidos pSAE4 y PBB5 fueron determinadas según lo descrito

en la Fig. 2. del artículo "Determination of the half-life of gdhA mRNA" de Merino et al., 1987a.

IV. RESULTADOS

4.1. Actividad específica de la enzima GDH y

estabilidad del mRNA de gdhA:

Stern et al. 1984 (74), reportaron que deleciones de las secuencias REP del extremo terminal del gene hisJ de Salmonella typhimurium, origina un decremento en la actividad específica de su producto en aproximadamente un 50%. En nuestro laboratorio hemos comprobado que ocurre un decremento similar en la actividad específica de la enzima GDH cuando una de las dos secuencias REP originales es removida del extremo terminal 3' del gene gdhA.

Con el objeto de estudiar la causa de tal decremento, la cepa de Escherichia coli PA340 auxótrofa para glutamato, fué independientemente transformada con los plásmidos pSAE4 y pBB5. El primero de estos plásmidos posee el gene gdhA nativo y sus regiones flanqueadoras incluyendo a las dos secuencias REP en el extremo 3' terminal. El plásmido pBB5 fué construido a partir del plásmido pSAE4 mediante la deleción de la segunda de las secuencias REP (ver Fig. 3, Becerril et al., 1985; Fig. 1, Merino et al., 1987a).

Las cepas transformadas fueron crecidas en medio mínimo NN suplementado con 15 mM de NH_4Cl como fuente de nitrógeno hasta fase logarítmica tardía. De cada uno de los cultivos, fueron tomados los volúmenes necesarios para realizar en forma simultánea las determinaciones de la actividad específica de GDH y la vida media de sus transcritos correspondientes.

Los resultados de la determinación de la actividad específica se muestran en la tabla I de Becerril et al., 1985.

Con el objeto de abolir el inicio de transcripción, se adicionó a los cultivos rifampicina en una concentración final de 0.8 mg/ml. Posteriormente, fueron tomadas alicuotas a intervalos de tiempo de 3 min, a las que se determinó la cantidad de mRNA específico de gdhA, tal y como se describe en el pie de la Fig. 2 del artículo de Merino et al., 1987a. En esa figura se muestra uno de los autoradiogramas empleados en dicha cuantificación. La cinética de degradación del mRNA de gdhA transcritos de los plásmidos pSAE4 y pBBS, se muestra en la Fig. 3 del mismo artículo.

4.2 Búsqueda por computadora de secuencias REP:

Con el objeto de localizar a las secuencias REP, en aquellos genes cuya secuencia nucleotídica ha sido reportada, se elaboró el programa de computadora "Localiza.REPS", cuyo listado se anexa en el apéndice I de esta tesis. Las secuencias REP así encontradas, se compilan en las tablas I, II y III de Becerril et al., 1985 y en las figuras I y II de Merino y Bolívar, 1987c. Un estudio bibliográfico de cada uno de los genes encontrados en la búsqueda permitió conocer los aspectos relevantes a las secuencias REP concernientes a dichos genes. Esta información se incluye en los pies de figuras y tablas antes mencionados. Es importante mencionar que a pesar de que la búsqueda de las secuencias REP se realizó en todas las secuencias de organismos procariotes y eucariotes compiladas en la base de datos GenBank,

las secuencias REP encontradas en el estudio pertenecen exclusivamente a los genomas de Escherichia coli y Salmonella typhimurium.

La información generada en esta búsqueda permitió conocer los diferentes operones que poseen secuencias REP. Algunos de estos operones se han secuenciado tanto en Escherichia coli como en otros organismos. El estudio comparativo de las secuencias nucleotídicas de los operones rpsU-dnaG-rpoD, del operón de biosíntesis de metionina; así como del operón glnALG, mostro que las secuencias REP del gene de un organismo de una especie, no se encuentran necesariamente en el gene homólogo del organismo de la otra especie, aún y cuando dichos genes posean entre sí una similitud global muy grande. El resultado de estas comparaciones se muestra en las figuras 1, 2 y 3. Durante el desarrollo de esta tesis, fueron publicados estudios comparativos similares a los anteriormente mencionados, que concuerdan con los resultados mencionados (59, 23, 51).

Cuando el grado de extríngecia[?] de la búsqueda fué reducido, con el programa "Localiza.REPs" se encontraron secuencias muy parecidas al consenso de las secuencias REP a las cuales se les denominó secuencias tipo-REP. Aún y cuando la similitud con el consenso de las secuencias tipo-REP es menor, permitieron encontrar semejanzas interesantes con otros locus bacterianos, como lo son los rRNA 16S y 23S de Escherichia coli, el ile-tRNA de Escherichia coli y Bacillus subtilis (ver tabla IV de Becerril et al., 1985). También se encontró que las dos secuencias REP del gene gdhA pueden formar parte de una estructura semejante a la

del Glu-tRNA, tal y como se muestra en la Fig. 2 de Becerril et al., 1985. Finalmente, cabe mencionar que en algunos casos, la región que existe entre dos secuencias REP, es similar al sitio de pegado de la proteína regulatoria ntnC (2).

Con el fin de observar la distribución de las secuencias REP en el cromosoma de Escherichia coli, estas secuencias fueron ubicadas en el mapa del cromosoma de Escherichia coli que se muestra en la figura 4. Como puede observarse, no existe una distribución que obedezca a alguna regla fácilmente detectable.

4.3 Búsqueda por computadora de otro tipo de secuencias reiteradas

Con el fin de conocer si la secuencia REP es el único tipo de secuencia altamente reiterada dentro del genoma bacteriano, y en caso de no ser así, encontrar las similitudes y diferencias que otras posibles secuencias reiteradas pudieran compartir con REP, fueron desarrollados los programas "Elimina.redundancia", "Búsqueda.Hash" y "Localiza.cadenas", cuyos listados se anexan en el apéndice I de esta tesis. El diagrama de flujo seguido en el estudio, se muestra en la figura 5.

Las secuencia nucleotídicas analizadas en el estudio se encuentran compiladas en la base de datos GenBank, versión 44.0, corresponde a agosto de 1986 (84). En ocasiones, mas de una entrada en esta base de datos, puede hacer referencia al mismo locus, tal como es el caso de la publicación simultánea de una secuencias por dos autores diferentes. A continuación se ejemplifica tal duplicación:

ECOGDHA E. coli gdhA gene coding for NADP-specific glutamate

dehydrogenase. 1179bp

ECOGDHA E. coli gdhA gene encoding for NADP-dependent glutamate dehydrogenase. Complete cds. 1902bp

Como esta reiteración de la información modifica la frecuencia real de aparición de secuencias que existen en el genoma, se elaboró el programa "Elimina.redundancia" que permite, dada una secuencia o fracción de ella, encontrar en su vecindad aquellas otras secuencias que posean un grado de similitud preestablecido. La restricción de solo buscar secuencias similares localmente, hizo posible reducir el tiempo de procesador central de la computadora en varios órdenes de magnitud. El resultado de este programa permitió identificar:

a) Secuencias de genes reportadas mas de una vez. Estas secuencias fueron ~~el~~ eliminadas del GenBank para construir un banco de información no redundante.

b) Secuencias de genes que se encuentran realmente duplicadas en el cromosoma. Tales son los casos de algunos genes de operones de tRNA y operones de genes ribosomales.

c) Secuencias de genes que poseen una duplicación interna. Tal es el caso del gene de la ribonucleósido difosfato reductasa nrda de Escherichia coli cuyo extremo 3' terminal posee una posible duplicación que incluye a dos secuencias REP. La primera de estas secuencias se encuentra adentro de la región codificadora del gene. Estos resultados se describen en el artículo de Merino y Bolivar, 1987b, que se incluye en el apéndice II de esta tesis. Este es el primer reporte de una

secuencia REP localizada dentro de un gene estructural, así como el primer ejemplo de secuencias REP aparentemente generadas por una duplicación genómica.

Una vez que la base de datos fué depurada, se utilizó el programa "Búsqueda.Hash" (ver apéndice I) para encontrar aquellas secuencias con altas frecuencia de reiteración. De acuerdo a una función "f" preestablecida, este programa permite descomponer a una secuencia de longitud determinada llamada cadena, en uno y solo un número. Este número representa la posición "P" de un elemento dentro de un vector, en el que se contabiliza la aparición de cada cadena decodificada. La longitud "L" de las cadenas a decodificar, depende de la capacidad del procesador de la computadora utilizada, así como al número de elementos diferentes "E" con la que se representa a la secuencia (E=4 para secuencias de DNA o RNA y E=20 para secuencias de aminoácidos). Debido a las anteriores restricciones, el tamaño de las cadenas analizadas correspondió a un dodecámero.

Una vez que conocidas las posiciones de los elementos del vector con mayor valor, se aplicó la función inversa de "f" para obtener la secuencia de las cadenas de mayor frecuencia de aparición. Estas cadenas de doce bases constituyeron la unidad mínima de búsqueda de elementos reiterados de mayor longitud. En ocasiones fué posible sobreponer la secuencia de dos o mas de ellos, dando origen a secuencias de mayor tamaño. Estas secuencias fueron posteriormente buscadas en el banco de datos GenBank, haciendo uso del programa "Localiza.cadenas". Los resultados obtenidos fueron los siguientes:

a) Los programas "Elimina.redundancia", "Búsqueda.Hash" y "Localiza.cadenas", fueron elaborados para encontrar cualquier secuencia altamente reiterada y con ellos fué posible localizar la mayoría de las secuencias REP previamente encontradas con el programa de búsqueda específica "Localiza.REPs".

b) No se encontró ninguna otra secuencia cuya frecuencia de aparición fuera tan alta como el de las secuencias REP. Sin embargo, fué posible encontrar a una familia de secuencias palindrómicas extragénicas constituida por terminadores rho-independientes bidireccionales que se muestra en la figura 3.

Figura 5. Diagrama de flujo para la búsqueda de secuencias reiteradas en banco de datos.

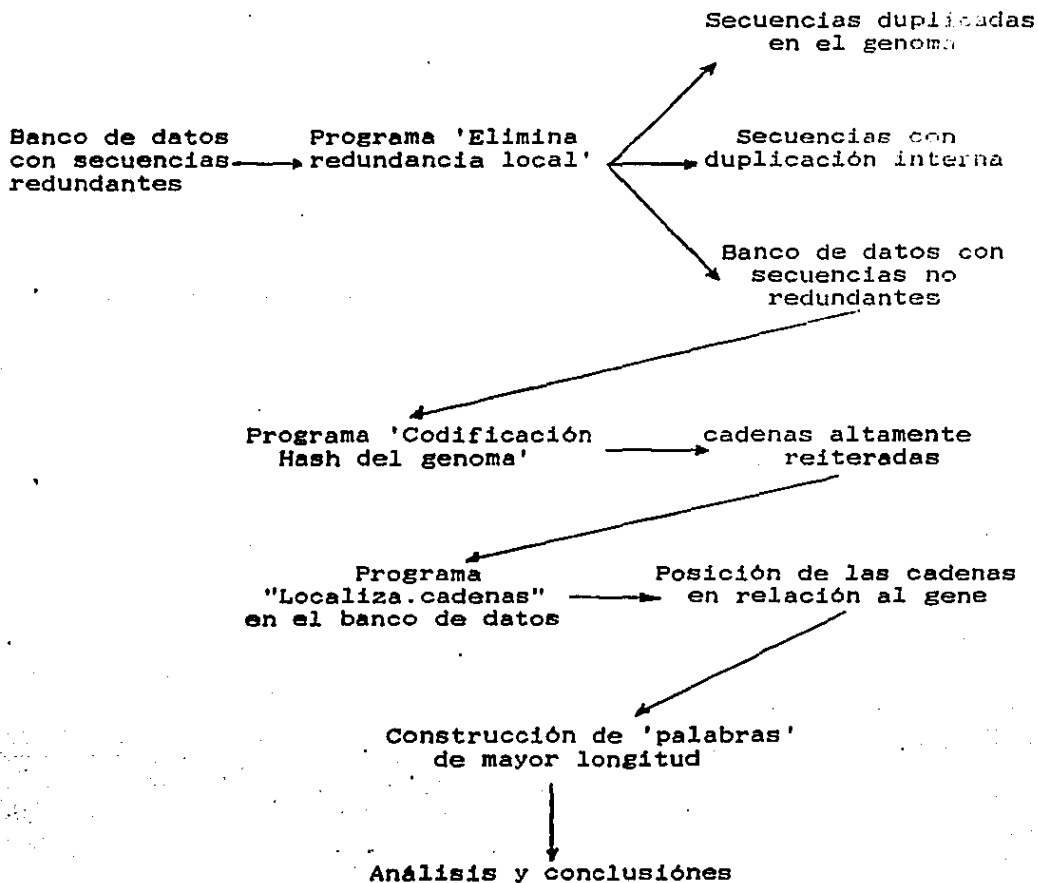


Figura 6. Familia de terminadores rho-independientes bidireccionales

- 1.- Consenso
- 2.- rpmG (proteína ribosomal L33)
- 3.- rplM (proteína ribosomal S9)
- 4.- rpsT (proteína ribosomal S20)
- 5.- sulA (supresor Lon)
- 6.- rpoC (subunidad beta de la RNA polimerasa)
- 7.- tap (proteína aceptora de metilo)
- 8.- Péptido líder del operón de leucina
- 9.- Péptido líder del operón de treonina
- 10.- Péptido líder del operón triptofano
- 11.- Péptido líder del operón ilvGEDA
- 12.- Péptido líder del operón de varios tRNAs
- 13.- Péptido líder del operón ilvB
- 14.- Met-tRNA-f2 (metY)

1.-	<u>AAAAAACCCCGCC</u>		<u>GGCGGGGTTTTTTT</u>
2.-	<u>taacAAAAAACCCCGCC</u>	cc	<u>GGCGGGGTTTTTTgTta</u>
3.-	<u>AAAAAACCC GC</u>	ttc	<u>GGC GGGTTTTTTTa</u>
4.-	<u>AAAAAGCCC GC</u>	gcaa	<u>GC GGGTTTTTTTaT</u>
5.-	<u>gAAAAAC GC</u>	tgct	<u>GC GGGTTTTTTTTT</u>
6.-	<u>tAAAAACCC GC</u>	ttcg	<u>GC GGGTTTTTTTaT</u>
7.-	<u>AAtEAACCC G atE</u>	atctgg cat	<u>C GGGTTTTTTTaTTT</u>
8.-	<u>AAAAAACCC GCgE</u>	catt	<u>gCG GGGTTTTTTTaT</u>
9.-	<u>AAAAAGCCC GCac</u>	ctgaca	<u>gtCGGGGcTTTTTTT</u>
10.-	<u>tAcccAgCCCCCct</u>	aatg	<u>agCGGGcTTTTTTTTg</u>
11.-	<u>tAAgACCCCCGcacc</u>	gaaa ggt	<u>CgGGGGGTTTTTTTTg</u>
12.-	<u>tttAAACCCCGC</u>	ttc	<u>GGCGGGGTTTTTTgTT</u>
13.-	<u>ccAAAAACCCCGCCg</u>	gCGcaaccg	<u>GGCGGGGTTTTTcgTT</u>
14.-	<u>AtAAAACCCCGat</u>	ttat	<u>CGGGGTTTTTTgTT</u>

Familia de terminadores rho-independientes bidireccionales de Escherichia coli encontrados en la búsqueda por computadora de secuencias altamente reiteradas. Las bases idénticas al consenso se muestran en mayúsculas. Las bases complementarias a cada uno de los lados del palíndrome se encuentran subrayadas.

V. DISCUSION

Análisis recientes del genoma procarionte revelaron la existencia de las secuencias REP. La estructura palindrómica de estas secuencias, el alto grado de conservación entre ellas, así como el gran número de veces en que se encuentran reiteradas, hacen de REP un elemento sin precedentes en el estudio del genoma bacteriano. Una de las propiedades inicialmente observadas de esta secuencia, fué su capacidad de alterar la concentración del producto final de los genes en los que se encuentra. Esta capacidad de las secuencias REP, aunado al hecho de encontrarla ampliamente distribuida a lo largo de todo el genoma, permitió proponerlas como un elemento novedoso de la regulación génica. Al inicio del presente estudio (octubre de 1985) no se conocía con exactitud los elementos involucrados en esta posible regulación.

En nuestro laboratorio hemos reportado que la delección de una secuencia REP del extremo terminal 3' del gene gdhA de Escherichia coli, origina una disminución de aproximadamente un 50% en la actividad específica de la enzima GDH (Becerril et al., 1985). Con el objeto de identificar el nivel en el que se realizaba dicho decremento, fueron realizados estudios de estabilidad de RNA mensajero de gdhA. Los resultados obtenidos en dicho estudio, revelaron que los transcritos de gdhA que poseen dos secuencias REP son aproximadamente dos veces más estables que aquellos que sólo poseen una de estas secuencias. Estos

resultados de estabilidad concuerdan con los recientemente obtenidos por Newbury et al., en el operón de transporte de histidina de Salmonella typhimurium. (59).

La regulación de la estabilidad del mRNA pueda ser un elemento importante en la expresión génica. Nilsson et al. (60), han reportado que existen transcritos cuyas vidas medias varían dependiendo de la velocidad de crecimiento (ompA y cat), mientras que la estabilidad de otros permanece inalterada (llp y bla). Resultados semejantes se han recientemente reportado en el cambio de estabilidad de transcritos en Klebsiella pneumoniae, que dependen de la concentración de la disponibilidad de las fuentes de nitrógeno en el medio así como de la cantidad de oxígeno disuelto en el medio (18).

No obstante, es conveniente mencionar que la estabilidad del mRNA no es el único factor limitante en la biosíntesis de una proteína, y que a pesar de que un transcrito pueda ser estabilizado de su degradación 3'-->5', no se excluye la posibilidad de que éste sea inactivado en su extremo 5' terminal. Aún y cuando no se ha demostrado conclusivamente la presencia de exonucleasas 5'-->3' en Escherichia coli (12), existen diferentes razones que permiten pensar que tal dirección de degradación exonucleolítica esté presente en organismos procariotes, o que la inactivación endonucleolítica en el extremo 5' terminal se lleve a cabo. Como se mencionó en la sección "Estabilidad del mRNA bacteriano" del capítulo II, una gran porción de los transcritos en organismos procariotes, se encuentran en proceso de decaimiento. Sería metabólicamente poco

eficiente que la célula sintetizara proteínas incompletas de transcritos que hayan sido degradados a partir del extremo 3'. De forma contraria, transcritos degradados a partir del extremo 5', carecerían de las señales indispensables para el inicio de traducción. El hecho de que algunos transcritos de fagos sean degradados e inactivados a partir del extremo 3' (50, 31, 71, 33), puede deberse a la necesidad de una respuesta rápida para inactivar proteínas que intervienen en decisiones tales como la lisis o lisogenia. No obstante todo lo anterior, es importante señalar que se han reportado transcritos de Escherichia coli, como el de ompA (79) y el del mensajero polisistrónico de triptofano (58), que decaen a partir de su extremo terminal 3'.

Las secuencias REP no son los únicos elementos capaces de estabilizar al mRNA. Actualmente se sabe que estructuras de tipo tallo-asa ubicadas en el extremo 3' o 5' de un transcrito, poseen la capacidad de estabilizarlo (6, 31, 33, 50, 58, 71, 81, 30). No obstante, ninguno de estos elementos posee una secuencia tan conservada, y en todos los casos sus tamaños son en mucho inferiores a REP. Por lo anterior, es posible pensar que el efecto estabilizador de las secuencias REP en el mRNA, no es necesariamente la única función de estas secuencias. Sin embargo esta propiedad pudo haber sido seleccionada en el organismo para desempeñar alguna función metabólicamente importante.

Por otro lado, es conveniente mencionar que el estudio comparativo de las secuencias nucleotídicas de operones de organismos de diferente especie, mostró que las secuencias REP de un organismo (ver figuras 1, 2 y 3), no necesariamente se

encuentra en el gene homólogo de otro organismo. Aún más, pese a que entre dichos genes existan regiones de alta similitud al rededor de REP, sólo diferirán específicamente en la región ocupada por la secuencia REP. Tal es el caso de la región intercistrónica glnA-glnLG de Escherichia coli, Salmonella typhimurium y Klebsiella pneumoniae (figura 3). Se ha descrito que en esta región intercistrónica, existen varios elementos indispensables para el control del operón que incluyen un posible terminador rho-independiente (47, 76, 51), un promotor para glnLG, así como su correspondiente sitio Shine-Dalgarno (76,51), y el sitio de unión del represor codificado por glnG (21, 76, 52, 29). Tal y como lo indican MacFarlane y Merrick (51), el hecho de que el operón glnALG se regule de forma muy parecida en estos tres organismos, pero que sin embargo, sólo el primero de ellos posea un par de secuencias REP, permite pensar de que al menos en este caso, REP no posee un papel determinante en el proceso de regulación. Conclusiones similares pueden obtenerse al analizar comparativamente las secuencias nucleotídicas de Escherichia coli y Salmonella typhimurium del operón rpsU-dnaG-rpoD y del operón de biosíntesis de metionina.

El estudio por computadora de secuencias altamente repetidas mostró que en el genoma de Escherichia coli no existe ninguna otra secuencia cuya frecuencia de aparición o grado de conservación sean tan altos como el de la secuencia REP. No obstante, fué posible detectar la existencia de una familia de terminadores rho-independientes bidireccionales en la región líder de algunos operones biosintéticos, así como en la región 5'

de algunos operones que codifican para proteínas ribosomales (ver figura 6). Los elementos de esta familia comparten con las secuencias REP el ser palindrómicos y extragénicos. Es posible pensar que la propiedad de palindromia, en ambos casos esté asociada con la capacidad de formar estructuras secundarias estables, ya que como se mencionó en el capítulo I, no obstante algunos elementos REP puedan diferir de la secuencia consenso de REP en algunas de sus bases, la capacidad de apareamiento entre los brazos de sus palindromes se conserva.

Stern et al., han propuesto que las secuencias REP pudieran haber sido originadas a partir de algún tipo de secuencia de inserción IS (74). Es interesante notar que, pese a que en el estudio de computadora no se identificaron a las secuencias de inserción IS como secuencias altamente reiteradas, estos elementos comparten con las secuencias REP algunas propiedades importantes: a) Al igual que las secuencias REP y la familia de terminadores rho-independientes bidireccionales antes mencionados, los extremos (también llamados brazos) de las secuencias de inserción IS, son secuencias repetidas invertidas. b) Se sabe que las secuencias IS son elementos capaces de proporcionar al organismo diversidad en su genoma ya que al transponerse pueden causar deleciones, inversiones, inserciones o translocaciones de fragmentos de DNA. Dado el alto grado de similitud entre las secuencias REP, se ha propuesto que éstas pudieran estar involucradas en eventos de recombinación dando origen a cierto tipo de rearrreglos cromosomales. c) Tanto las secuencias REP como las secuencias IS entre Escherichia coli y de

Salmonella typhimurium, no poseen posiciones equivalentes en sus mapas genómicos. (59, 23, 51,20). d) Tanto las secuencias de inserción IS como las secuencias REP se han propuesto como tipos de "DNA egoísta" en el genoma bacteriano (20,74). No obstante las anteriores similitudes entre estas dos secuencias, es claro que la frecuencia de aparición de las secuencias REP es varias veces superior al de las secuencia de inserción IS, y que dada su diferencia en tamaño, organización y función, es difícil elaborar hipótesis verificables sobre el origen de estos dos elementos.

A pesar de que diferentes programas con algoritmos de almacenamiento "Hash-coding" de computadora han sido desarrollados para el análisis de secuencias de tipo biológico (22, 37, 38, 39, 40, 24), los resultados del análisis de secuencias reiteradas anteriormente expuestos son, hasta lo que nosotros sabemos, los primeros de este tipo que se reportan en la literatura. Debido al gran requerimiento de memoria que los programas de búsqueda Hash utilizan, sólo se han podido utilizar con éxito en genomas de tamaño pequeño, como son los de bacteriofagos y virus. Posiblemente esta sea la razón por la que, a pesar de que algunos de estos programas fueran desarrollados antes de la primera compilación de las secuencias REP por búsqueda directa en el año de 1984 (74, 25), no fueran capaces de detectar la existencia de esta secuencia reiterada o alguna otra, como la de la familia de terminadores rho-independientes bidireccionales antes mencionada. No obstante, la estrategia de encontrar secuencias de gran longitud a partir de unidades de

menor tamaño que facilitó el programa "Busqueda.Hash" desarrollado en esta tesis, que permitió encontrar la mayoría de las secuencias REP previamente localizadas con el programa de búsqueda directa "Localiza.REPs".

Es importante notar que los programas "Elimina.redundancia", "Búsqueda.Hash" y "Localiza.cadenas" fueron elaborados con el objeto de encontrar cualquier secuencia altamente reiterada, y que el hecho de haber encontrado a las secuencias REP, constituyó un control positivo de la capacidad de detección de los mismos.

La información redundante en bancos de datos de secuencias de ácidos nucleicos y proteínas, dificulta el estudio comparativo de la información que en ellos se encuentra. En la actualidad, son pocas las bases de datos que han fijado su atención en resolver este problema. Un ejemplo de este tipo de base de datos es el recientemente elaborado banco de secuencias de proteínas PseqIP, del Instituto Pasteur, Francia (15). Debido al hecho de carecer una base similar de secuencias de ácidos nucleicos, fue necesario elaborar un banco de datos que contuviera las secuencias de los genes de Escherichia coli en forma no redundante. Los resultados obtenidos en la elaboración de esta base demuestran que el algoritmo no exhaustivo del programa "Búsqueda.local", permite eliminar una gran mayoría de la información redundante.

Una de las características mas importantes de la secuencias REP, y que de hecho da origen a su nombre, es la de estar localizadas en las regiones extragénicas de los operones. No obstante, el estudio de las secuencias altamente reiteradas en el

genoma de Escherichia coli, permitió encontrar a una secuencia REP en la región extructural del gene ribonucleósido difosfato reductasa nrda (ver apéndice II). Este es el primer, y hasta el momento, el único reporte de una secuencia REP en la región que codifica para una proteína.

El hecho de encontrar comunmente a REP en regiones extragénicas, puede obedecer a diferentes causas. Posiblemente, la presión de selección a la que está sujeta cualquier secuencia nucleotídica que codifica para una proteína, es una de las principales razones que explica la imposibilidad de que una secuencia altamente conservada se encuentre presente en un número muy grande de genes estructurales cuyos productos proteicos, en la mayoría de los casos, no poseen ninguna relación estructural o funcional entre sí. Lo anterior no excluye la posibilidad de que, en un momento determinado, alguna secuencia REP pudiera formar parte de la región estructural de un gene, tal y como sucede con la primera de las cinco secuencias REP del gene nrda. Es interesante observar que esta secuencia REP, posee un menor grado de similitud con el consenso de REP, que cualquiera de las otras cuatro secuencias (ver apéndice II, figura del artículo: "The ribonucleoside diphosphate reductase (nrda) gene of Escherichia coli carries a Repetitive Extragenic Palindromic (REP) sequence in it's 3' structural terminus"). Muy probablemente, cualquier elemento REP dentro de una región codificatante sujeta a presión selectiva, divergirá del consenso de las secuencias REP con mayor velocidad que cualquier otra secuencia REP libre de presiones de esta

índole, de tal forma que las secuencias REP casi siempre se localizarán extragénicamente o bien no será estadísticamente posible diferenciarlas de cualesquier otras secuencias azarosamente similares a REP.

Mediante el programa de búsqueda directa "Localiza.REPs", fué posible encontrar 82 secuencias REP, 23 de las cuales no habían sido previamente reportadas. La determinación de la posición de estas secuencias en el mapa del cromosoma de Escherichia coli (figura 4), permite observar que estos elementos se encuentran dispersos en todo el genoma, y que tal y como lo habían previamente señalado Stern et al., su orientación, posición y distribución de las secuencias REP, parece no obedecer a ninguna regla conocida (74).

Es importante mencionar que el estudio del gene nrda de Escherichia coli, que se incluye en el apéndice II de esta tesis, constituye el único ejemplo reportado de un gene que posee cinco secuencias REP. El análisis de la secuencia nucleotídica de este gene, reveló que muy probablemente, dos de estas cinco secuencias REP, fueron originadas a partir de una duplicación génica. Newbury et al. (59), han recientemente propuesto a las secuencias REP como un nuevo tipo de "DNA-egoísta" en el genoma procarionte, y postulan que REP se ha dispersado por todo el genoma mediante un mecanismo de "conversión génica mediado a nivel de mRNA". Aún y cuando los autores no describen con claridad este proceso, es posible pensar que en caso de que éste existiera, no sería el único mecanismo

involucrado en la generación de las secuencias REP, ya que eventos de duplicación génica, tal como el posiblemente acaecido en nrpA, pueden estar involucrados en tal tarea.

VIII. BIBLIOGRAFIA

- 1.- Ames, B.W., Tsang, T.H., Buck, M. and Christman, M.F.: The leader mRNA of the histidine attenuator region resembles tRNA-His: Possible regulatory implications. Proc. Natl. Acad. Sci. USA 80 (1983) 5240-5242.
- 2.- Ames, G.F.-L., and Nikaido, K.: Nitrogen regulation in Salmonella typhimurium. Identification of ntnC protein-binding site and definition of a consensus binding sequence. EMBO J. 4 (1985) 539-547.
- 3.- Avery, O., McLead, C.M. and McCarthy, M.J.: Studies on the chemical nature of the substance inducing transformation of Pneumococcal types. Exper. Med., 79 (1944) 137-158.
- 4.- Becerril, B., Valle, F., Merino, E., Riba, L. and Bolivar, F.: Repetitive extragenic palindromic (REP) sequences in the Escherichia coli gdhA gene. Gene 37 (1985) 53-62.
- 5.- Becerril, B.: Estudio y caracterización de las regiones regulatorias del gene que codifica para la enzima glutamato deshidrogenasa de Escherichia coli K-12. Tesis Doctoral. UNAM. 1986.
- 6.- Belasco, J.G., Beatty, J.T., Adams, C.W., von Gabain, A. and Cohen, S.N.: Differential expression of photosynthesis genes in E. capsulata results from segmental differences in stability within the polycistronic rxCA transcript. Cell 40 (1985) 171-181.
- 7.- Belasco, J.G., Nilsson, G., von Gabain, A. and Cohen, S.N.: The stability of E. coli gene transcripts is dependent on determinants localized to specific mRNA segments. Cell 46 (1986) 245-251.
- 8.- Blundell, M., Craig, E. and Kennell, D.: Decay Rates of Different mRNA in E. coli and models of Decay. Nature New Biol. 238 (1972) 46-49.
- 9.- Bolivar, F. and Backman, K. En Methods in Enzymology (Wu, R., eds.) Academic Press, New York, NY. Vol. 68 (1979) pp. 245-268.
- 10.- Bolton et al., Carnegie Inst. Wash. Year Book 62, 303 (1963)
- 11.- Bolton et al., Carnegie Inst. Wash. Year Book 63, 309 (1964)
- 12.- Bothwell, A.L.M. and Apirion, D.: Biochem. Biophys. Res. Commun. 44 (1971) 844-851.

- 13.- Bradford, M.: A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Analyt. Biochem.* 72 (1976) 248-254.
- 14.- Britten, R.J. and Kohene, D.E.: Repeated sequences in DNA. *Science* 161 (1968) 529-540.
- 15.- Claverie, J.M. and Bricault, L.: PseqIP: A Nonredundant and Exhaustive Protein Sequence Data Bank Generated From 4 Major Existing Collections. *PROTEINS: Structure, Function, and Genetics.* 1 (1986) 60-65.
- 16.- Clewell, D.: Nature of ColE1 plasmid recombination in the presence of chloramphenicol. *J. Bacteriol.* 110 (1972) 667-676.
- 17.- Cohen, S.N., Chang, A.C.Y. and HSU, L.: Genetic transformation of Escherichia coli by R factor DNA. *Proc. Natl. Acad. Sci. USA* 69 (1972) 2110-2114.
- 18.- Collins, J.J., Roberts, G.P. and Winston, J.B.: Posttranscriptional Control of Klebsiella pneumoniae nif mRNA Stability by the nifL Product. *J. Bacteriol.* 168 (1986) 173-178.
- 19.- Covarrubias, A.A., Sanchez-Pescador, R., Osorio, A., Bolivar, F. and Bastarrachea, F.: ColE1 hybrid plasmids containing Escherichia coli genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* 3 (1980) 150-164.
- 20.- Deonier, R.C.: Location of Native Insertion Sequence Elements. En Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology. (Ingraham, J.L., Brooks, K., Magasanik, B., Schaechter, M. and Umberger, E., Eds.) American Society for Microbiology. Washington, D.C. vol 2. (1987) pp. 982-989.
- 21.- Dixon, R.: Tandem promoters determine regulation of the Klebsiella Pneumoniae glutamine synthetase (glnA) gene. *Nucl. Acids. Res.* 12 (1984) 7811-7830.
- 22.- Dumas, J.P. and Ninio, J.: Efficient algorithms for folding and comparing nucleic acids sequences. *Nucleic Acids Res.* 10 (1982) 197-206.
- 23.- Erickson, B.D., Burton, Z.F., Watanabe, K.K. and Burges, R.R.: Nucleotide sequence of the rpsu-dnaG-rpoD operon from Samonella typhimurium and a comparison of this sequence with the homologous operon of Escherichia coli. *Gene* (1985) 67-78.
- 24.- Fondarat, C., Dessen, P. and Le Beux, F.: Principle of codification for quick comparisons with the entire biomolecule databanks and associated programs in FORTRAN 77. *Nucleic Acids Research.* 14 (1986) 197-204.

- 25.- Gilson, E., Clément, J.-M., Brutlag, D. and Hofnung, M.: A family of dispersed repetitive extragenic palindromic DNA sequences in E. coli. EMBO J. 3 (1984) 1417-1421.
- 26.- Gilson, E., Ferrin, D., Clément, J.M., Szmelcman, S., Dassa, E., Hofnung, M.: Palindromic units (P.U.s) from E. coli as binding sites for a chromoid associated protein. FEBS Lett. 206 (1986) 323-328.
- 27.- Gilson, E., Clément, J.M., Ferrin, D. and Hofnung, M.: Palindromic units: a case of highly repetitive DNA sequences in bacteria. Trends Genet. 3 (1987) 226-230.
- 35.- Gilson, E., Ferrin, D., Saurin, W. and Hofnung, M.: Species Specificity Palindromic Units. J. Mol. Evol. 25 (1987) 371-373.
- 29.- Ginsburg, A.: Glutamine synthetase of Escherichia coli: some physical and chemical properties. Adv. Protein Chem., 26 (1982) 1-79.
- 30.- Gorski, K., Roch, J.-M., Prentki, P. and Krisch, H.M.: The stability of bacteriophage T4 gene 32 mRNA: a 5' leader sequence that can stabilize mRNA transcripts. Cell 43 (1985) 461-469.
- 31.- Guarneros, G., Montanez, C., Hernandez, T. and Court, D.: Posttranscriptional control of bacteriophage int gene expression from a site distal to the gene. Proc. Natl. Acad. Sci. USA 79 (1982) 238-242.
- 32.- Haemi, A., Joshi, S. and Champeville, F., in CoHN, W. E. (Ed.), Progress in Nucleic Acids Research and Molecular Biology. Academic Press, New York, 1982, pp. 85-102.
- 33.- Hayashi, M.N. and Hayashi, M.: Cloned DNA sequences that determine mRNA stability of bacteriophage X174 in vivo are functional. Nucl. Acids Res. 13 (1985) 5937-5948.
- 34.- Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clément, J.M. and Hofnung, M.: A novel intercistronic regulatory element of prokaryotic operons. Nature 298 (1982) 760-762.
- 35.- Jacquet, M. and Kepes, A.: Initiation, Elongation and Inactivation of lac messenger RNA in Escherichia coli studied by measurement of its beta-Galactosidase synthesis capacity in vivo. J. Mol. Biol. 60 (1971) 453-472.
- 36.- Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Ramaut, E. and Helinski, D. (1979) in Methods in Enzymology (Wu, R., eds.) Vol. 68, pp. 268-280, Academic Press, New York, NY.
- 37.- Karlin, S., Ghandour, G., Ost, F., Tavaré, S. and Korn, L.J.: New approaches for computer analysis of nucleic acids sequences. Proc. Natl. Acad. Sci. USA 80 (1983) 5660-5664.

38.- Karlin, S. and Ghandour, G.; Comparative statistics for DNA and protein sequences: Single sequence analysis. Proc. Natl. Acad. Sci. USA 82 (1985) 5800-5804.

39.- Karlin, S. and Ghandour, G.; Comparative statistics for DNA and protein sequences: Multiple sequence analysis. Proc. Natl. Acad. Sci. USA 82 (1985) 6180-6190.

40.- Karlin, S.; Significant potential secondary structures in the Epstein-Barr virus genome. Proc. Natl. Acad. Sci. USA 83 (1986) 6915-6919.

41.- Kennell, D. and Bicknell.; Decay of messenger ribonucleic acids from the lactose operon of Escherichia coli as function of growth temperature. J. Mol. Biol. 74 (1973) 21-31.

42.- Kennell, D. and Riezman, H.; Transcription and Translation Initiation Frequencies of the Escherichia coli lac operon. J. Mol. Biol. 114 (1977) 1-21.

43.- Kennell, D.; The instability of messenger RNA in bacteria. en Maximizing Gene Expression. Reznikoff, W. and Gold, L. eds. Butterworths Boston, 1986.

44.- Kepes, A.; Kinetics of induced enzyme synthesis. Determination of the mean life of galactosidase specific-messenger RNA. Biochim. Biophys. Acta 76 (1963) 293-309

45.- Kepes, A. and Begun, S.; Peptide chain initiation and growth in induced synthesis of beta-galactosidase. Biochim. Biophys. Acta 123 (1966) 546-560.

46.- Kepes, A.; Sequential transcription and translation in the lactose operon of Escherichia coli. Biochim. Biophys. Acta 138 (1967) 107-123.

47.- Krajewska-Gryniewicz, K. and Kustu, S.; Mol. Gen. Genet. 193 (1984) 135-142.

48.- Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H.; RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 16 (1977) 4743-4751.

49.- Littauer, U.Z. (1982) Prog. Nucl. Acids. Res. Mol. Biol. 27:53-83.

50.- Lozeron, H.A., Dahlberg, J.E. and Szybalski, W.; Processing of the major leftward mRNA of coliphage lambda. Virology 71 (1976) 262-277.

- 51.- MacFarlane, S.A. and Merrick, M.; The nucleotide sequence of the nitrogen regulation gene ntrB and the glnA-ntrBC intergenic region of Klebsiella pneumoniae. Nucl. Acids Res. 13 (1985) 7591-7606.
- 52.- MacNeil, T., MacNeil, D. and Tyler, B.; Fine Structure Deletion Map and Complementation Analysis of the glnA-glnL-glnG region in Escherichia coli. J. Bacteriol. 150 (1982) 1302-1313.
- 53.- Maniatis, T., Fritsch, E.F. and Sambrook, J.; in Molecular cloning. A laboratory manual. Cold Spring Harbord Lab. 1982.
- 54.- Maxam, A.M. and Gilbert, W.; A new method for sequencing DNA. Proc. Acad. Sci. USA. 74 (1977) 560-564.
- 55.- Merino, E., Becerril, B., Valle, F. and Bolivar, F.; Deletion of a repetitive extragenic palindromic (REP) sequence downstream of the structural gene of Escherichia coli glutamate dehydrogenase affects the stability of its mRNA. Gene 58 (1987a) 311-314.
- 56.- Merino, E. y Bolivar, F.; The ribonucleoside diphosphate reductase (nrda) gene of Escherichia coli carries a Repetitive Extragenic Palindromic (REP) sequence in it's 3' structural terminus. Nucl. Acids Res. (1987b). En revisión.
- 57.- Merino, E. y Bolivar F.; A compilation of Repetitive Palindromic (REP) sequences. Nucl. Acids Res. (1987c). En revisión.
- 58.- Mott, J.E., Galloway, J.L. and Platt, T.; Maturation of Escherichia coli tryptophan operon mRNA: evidence for 3' exonucleolytic processing after rho-dependent termination. EMBO J. 4 (1985) 1887-1891.
- 59.- Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F.; Stabilization of Translationally Activr mRNA by Prokariotic REP sequences. Cell 48 (1987) 297-310.
- 60.- Nilsson, G., Belasco, J.G., Cohen, S.N.; Growth-rate dependent regulation of mRNA stability in Escherichia coli. Nature 312 (1984) 75-77.
- 61.- Pedersen, S. and Reeh, S.; Functional mRNA Half Lives in E. coli. Mol. Gen. Genet. 166 (1978) 329-336.
- 62.- Prince, J.B., Gutell, R.R. and Gatett, R.A.; A consensus model of the Escherichia coli ribosome. Trends Biochem. Sci. 8 (1983) 359-363.
- 63.- Riba, L.; Caracterización y estudio de la región flanqueadora 3' del gene que codifica para la enzima Glutamato Deshidrogenasa en Escherichia coli K-12. Tesis Profesional. UNAM. 1985.

- 64.- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P.: Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113 (1977) 237-251
- 65.- Rejinders, L., Sloof, P., Silva, J. and Borst, P.: Gel electrophoresis of RNA under denaturing conditions. *Biochim. Biophys. Acta* 324 (1973) 320-333.
- 66.- Roberts, R.J.: Restriction endonucleases. *Crit. Rev. Biochem.* 4 (1976) 123-164.
- 67.- Roberts, R.J.: The role of restriction endonucleases in Genetic Engineering; (1977) in *Recombinant Molecules: Impact on Science and Society*. Eds. Beers, R.F. Jr. and Bassett, E.G., Raven Press, New York, NY.
- 68.- Rodriguez, R.L. and Tait, R.C.: *Recombinant DNA techniques*. Addison-Wesley, Reading, MA., 1983, pp. 50-51.
- 69.- Rufus, M.W. and Julius, H.J.: Common Evolutionary Origin of the ilvGMEDA Attenuation Locus and Leu-tRNA-1 in Escherichia coli. *J. Bacteriol.* 169 (1987) 2893-2895.
- 70.- Sanger, F. and Coulson, A.: *J. Mol. Biol.*, 94 (1975) 441-448.
- 71.- Schmeissner, U., McKenney, K., Rosenberg, M. and Court, D.: Removal of a terminator structure by RNA processing regulates int gene expression. *J. Mol. Biol.* 176 (1984) 39-53.
- 72.- Smith, H. (1979) in *Methods in Enzymology* (Grossman, L. and Moldave, K., eds.) Vol. 65, pp. 560-580, Academic Press, New York, NY.
- 73.- Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98 (1975) 503-517.
- 74.- Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F.: Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37 (1984) 1015-1026.
- 75.- Thomas, White, R. and Davis, R.: Hybridization of RNA to double-strand DNA: Formation of R-loops. *Proc. Natl. Acad. Sci. USA.* 73 (1976) 2294-2298.
- 76.- Ueno-Nishio, S. Mango, S., Reitzer, L.J. and Magasanik, B.: Identification and Regulation of the glnL Operator-Promoter of the Complex glnALG Operon of Escherichia coli. *J. Bacteriol.* 160 (1984) 379-384.

77.- Valle, F., Sanvicente, E., Seeburg, P., Covarruvias, A., Rodriguez, R.L. and Bolivar, F.: Nucleotide sequence of the promoter and amino-terminal coding region of the glutamate dehydrogenase structural gene of Escherichia coli. Gene 23 (1983) 199-209.

78.- Valle, F., Becerril, B., Chen, E., Seeburg, P., Heyneker, H. and Bolivar, F.: Complete nucleotide sequence of the glutamate dehydrogenase gene from Escherichia coli K-12. Gene 27 (1984) 193-199.

79.- von Gabain, A., Belasco, J.G., Schottel, J.L., Chang, A.C.Y. and Cohen, S.N.: Decay of mRNA in Escherichia coli: Investigation of the fate of specific segments of transcripts. Proc. Natl. Acad. Sci. USA. 80 (1983) 653-657.

80.- Watson, J.D. and Crick, F.H.C.: Nature, 17 (1953) 737-738.

81.- Wong, H.C. and Chang, S.: Identification of a positive retroregulator that stabilizes mRNAs in bacteria. Proc. Natl. Acad. Sci. USA 83 (1986) 3233-3237.

82.- Worcel, A. and Burgi, E.: On the structure of the folded chromosome of Escherichia coli. J. Mol. Biol. 71 (1972) 127-147.

83.- Young, F.S. and Furano, A.V.: Regulation of the syntesis of E. coli elongation factor Tu. Cell 24 (1981) 695-706.

84.- The GenBank Genetic Sequence Data Bank: Floppy Diskette User's Guide. Release 44.0. August, 1986. National Institutes of Health. Department of Health and Human Services for the Genetic Sequence Data Bank. Cambridge, USA.

VII. PROGRAMAS DE COMPUTADORA DESARROLLADOS

7.1 Programa 'Localiza cadenas'

```
(*****)  
(* LOCALIZA CADENAS *)  
(* Este programa permite localizar, dada una secuencias *)  
(* especifica, a sus identicas o similares dentro del banco *)  
(* de datos de ácidos nucléicos GenBank. *)  
(*****)
```

type

```
INDEX_TYP=file of byte;  
BITE_TYP=array [1..8] of char;  
STR_TYP=string[80];
```

var

```
RESP: char;  
SEC: array [1..32760] of char;  
REP1,REP2: array[1..25,0..39] of char;  
WHY,SPACE:array[1..10] of char;  
CONR,MATCH: array[1..25] of integer;  
DREP: string[50];  
DIVISION: string[3];  
ENTRADA: string[10];  
BTA: array [1..10] of byte;  
ANNFILENUM,SEQFILENUM,BT1,BT2,BT3,  
FILEINFO,DIVNUM,TYPINFO,ECODE: byte;  
RECORDLENGTH,FIRSTLINE,LASTLINE,SEQRECNUM,  
MIN,MA,PP,NUM,PRINCIPIO,FIN,DUMMY,D1,D2,  
D3,D4,SCORE1,SCORE2,SCO,ANNRECNUM,nbus,  
SEGRECLENGTH,ANNRECLENGTH,AA,A,B,C,D,E: integer;  
LUG: array [1..200] of integer;  
INDEX,SEQUENCE: INDEX_TYP;  
BITARY: BITE_TYP;  
ADDRESS,BASENUM,FROM,TWO,BB: real;  
SALIDA: text;  
REP: boolean;
```

```
(*****)  
(* Decodifica secuencia *)  
(*****)
```

```
procedure PRINT(BITE:BITE_TYP;var DONE:boolean);  
begin
```

```
BASENUM:=BASENUM+1;  
NUM:=NUM+1;  
if BITE='00000001' then SEC[NUM]:='a'  
else if BITE='00000010' then SEC[NUM]:='c'  
else if BITE='00000100' then SEC[NUM]:='g'  
else if BITE='00001000' then SEC[NUM]:='t'  
else if BITE='00000101' then SEC[NUM]:='r'  
else if BITE='00001111' then SEC[NUM]:='n'
```

```

    else if BITE='00001010' then SEC[NUM]:= 'y'
    else if BITE='00000000' then DONE:=true;
    if BASENUM=32680 then DONE:=true;
    if DONE then begin
        NUM:=NUM-1;
        BASENUM:=BASENUM-1;
    end;
end;

```

```

{*****}
{*   Descompone la información leída en forma compacta para   *}
{*   para su posterior decodificación                          *}
{*****}

```

```

procedure INFLATE;

```

```

var

```

```

    BT:                               byte;
    A:                                 real;
    B,C,D:                             integer;
    BITARY1,BITARY2,BITARY3:          BITE_TYP;
    DONE:                               boolean;

```

```

begin

```

```

    BASENUM:=0;
    NUM:=0;
    A:=0;
    DONE:=false;
    repeat
        read(SEQUENCE,BT);
        A:=A+2;
        for B:=1 to 8 do begin
            BITARY1[B]:='0';
            BITARY2[B]:='0';
            BITARY3[B]:='0';
        end;
        for B:=8 downto 1 do begin
            C:=1;
            for D:=2 to B do C:=C*2;
            if BT-C>=0 then begin
                BITARY1[9-B]:='1';
                BT:=BT-C;
            end;
        end;
        for B:=1 to 4 do begin
            BITARY2[B+4]:=BITARY1[B];
            BITARY3[B+4]:=BITARY1[B+4];
        end;
        PRINT(BITARY2,DONE);
        PRINT(BITARY3,DONE);
    until DONE;
end;

```

```

{*****}
{*           Selecciona grupo de lectura           *}
{*****}

```

```

procedure GET_GROUP;
begin
  assign(INDEX,'G'+DIVISION+'.IDX');
  reset(INDEX);
  read(INDEX,BT1,BT2);
  RECORDLENGTH:=BT1+BT2*256;
  seek(INDEX,8);
  read(INDEX,BT1,BT2);
  FIRSTLINE:=BT1+BT2*256;
  read(INDEX,BT1,BT2);
  LASTLINE:=BT1+BT2*256;
  reset(INDEX);
end;

```

```

{*****}
{*           Calcula apuntadores para lectura           *}
{*****}

```

```

procedure FIND_POINTERS;
begin
  read(INDEX,FILEINFO,DIVNUM);
  ANNFILENUM:=FILEINFO div 16;
  SEQFILENUM:=FILEINFO-ANNFILENUM*16;
  read(INDEX,BT1,BT2);
  SEQRECNUM:=BT1+BT2*256;
  read(INDEX,BT1,BT2);
  ANNRECNUM:=BT1+BT2*256;
  assign(SEQUENCE,'G'+DIVISION+chr(SEQFILENUM+48)+'.SEQ');
  reset(SEQUENCE);
  read(SEQUENCE,BT1,BT2);
  SEQRECLENGTH:=BT1+BT2*256;
  seek(SEQUENCE,5);
  for A:=1 to 5 do begin
    read(SEQUENCE,BT1);
  end;
  reset(SEQUENCE);
  ADDRESS:=(SEQRECNUM/1-1.0)*(SEQRECLENGTH/1)+21.0;
  longseek(SEQUENCE,(ADDRESS-3));
  read(SEQUENCE,TYPINFO);
  longseek(SEQUENCE,ADDRESS);
end;

```

```

{*****}
{*          Cambia minúscula por mayúscula          *}
{*****}

```

```

procedure UPER;
begin
  case SEC[D1+D2] of
    'a': SEC[D1+D2]:='A';
    'c': SEC[D1+D2]:='C';
    'g': SEC[D1+D2]:='G';
    't': SEC[D1+D2]:='T';
  end;
end;

```

```

{*****}
{*  Compara secuencias y escribe resultados en caso de igualdad  *}
{*****}

```

```

procedure BUSQUEDA;
begin
  for D4:=1 to NBUS do begin
    write('buscando a: ');
    for D3:=0 to CONR[D4] do write(REP1[D4,D3]);
    writeln;
    REP:=false;
    D3:=0;
    PP:=0;
    for D1:=NUM+1 TO NUM+20 do SEC[D1]:='.';
    for D1:=1 to NUM do begin
      SCO:=0;
      for D2:=0 to CONR[D4] do
        if (REP1[D4,D2]=SEC[D1+D2]) or
            (REP2[D4,D2]=SEC[D1+D2]) then SCO:=SCO+1;
      if SCO>=MATCH[D4] then begin
        PP:=PP+1;
        LUG[PP]:=D1;
        REP:=true;
        D3:=D3+1;
        for D2:=0 to CONR[D4] do begin
          if REP1[D4,D2]=SEC[D1+D2] then UPER;
          write(SEC[D1+D2]);
          write(SALIDA,SEC[D1+D2]);
        end;
        writeln(SALIDA);
        writeln;
      end;
    end;
  end;
  if rep then begin
    writeln(ENTRADA,' ',D3);
    writeln(SALIDA,ENTRADA,' ',D3);
    for D1:=1 to PP do begin
      write(SALIDA,LUG[D1]:6,' ');
      MIN:=LUG[D1]-30;
    end;
  end;
end;

```



```

        if MIN<1 then MIN:=1;
        MA:=LUG[D1]+30;
        for D2:=MIN to MA do write(SALIDA,SEC[D2]);
        writeln(SALIDA);
    end;
    writeln(SALIDA);
    writeln;
end;
end;
end;
end;

```

```

{*****}
{*      Pregunta parametros de corrida      *}
{*****}

```

```

procedure INICIO;
begin
    for SCORE1:=1 to 50 do DREP[SCORE1]:=' ';
    NBUS:=0;
    RESP:='s';
    while (RESP='s') or (RESP='S') do begin
        write('Deseas buscar alguna secuencia? ');
        readln(RESP);
        if (RESP='s') or (RESP='S') then begin
            NBUS:=NBUS+1;
            for SCORE1:=1 to 50 do DREP[SCORE1]:=' ';
            write('Secuencia a buscar: ');
            readln(DREP);
            write('Score: ');
            readln(MATCH[NBUS]);
            SCORE1:=1;
            while (DREP[SCORE1]<>' ') do begin
                REP1[NBUS,SCORE1-1]:=DREP[SCORE1];
                case DREP[SCORE1] of
                    'a': REP2[NBUS,SCORE1-1]:='A';
                    'c': REP2[NBUS,SCORE1-1]:='C';
                    'g': REP2[NBUS,SCORE1-1]:='G';
                    't': REP2[NBUS,SCORE1-1]:='T';
                end;
                CONR[NBUS]:=SCORE1-1;
                SCORE1:=SCORE1+1;
            end;
        end;
    end;
end;
end;
write('Grupo ?
(PRI,ROD,MAM,VRT,INV,PLN,ORG,BCT,RNA,VRL,PHG,SYN,UNA) ');
readln(DIVISION);
write('Principio ? ( => ) ',FIRSTLINE,' ');
readln(PRINCIPIO);
write('Fin ? ( <= ) ',LASTLINE,' ');
readln(FIN);
end;

```

```

{*****}
{*          Programa Principal          *}
{*  Establece el orden de llamada de las diversas subrutinas  *}
{*****}

```

```
begin
```

```
INICIO;
```

```
  GET_GROUP;
```

```
  assign(SALIDA,'A:busca');
```

```
  rewrite(SALIDA);
```

```
  writeln(SALIDA,DIVISION);
```

```
  writeln(SALIDA);
```

```
  for dummy:=PRINCIPIO to FIN do begin
```

```
    ADDRESS:=((DUMMY-1)/1)*(RECORDLENGTH/1);
```

```
    longseek(INDEX,ADDRESS);
```

```
    for A:=1 to 10 do read(INDEX,BTA[A]);
```

```
    for A:=1 to 10 do WHY[A]:=chr(BTA[A]);
```

```
    ENTRADA:=WHY;
```

```
    writeln(DUMMY,' ',ENTRADA);
```

```
    for A:=1 to (10-length(ENTRADA)) do SPACE[A]:=' ';
```

```
    ENTRADA:=ENTRADA+SPACE;
```

```
    FIND_POINTERS;
```

```
    INFLATE;
```

```
    BUSQUEDA;
```

```
  end;
```

```
  close(SALIDA);
```

```
end.
```

7.2 Programa 'Localiza REPS'

```
{*****}
{*          LOCALIZA REPS          *}
{* Este programa localiza especificamente secuencias REP. *}
{* El algoritmo del programa esta basado en programa LOCALIZA *}
{* CADENAS descrito en el inciso 3.4.2. El listado del programa *}
{* solamente incluye aquellas subrutinas que son diferentes en *}
{* ambos programas. *}
{*****}
```

type

```
INDEX_TYP=file of byte;
BITE_TYP=array [1..8] of char;
STR_TYP=string[80];
```

var

```
RESP: char;
SEC: array [1..32760] of char;
REP1,REP2: array[1..25,0..39] of char;
WHY,SPACE:array[1..10] of char;
CONR,MATCH: array[1..25] of integer;
DREP: string[50];
DIVISION: string[3];
ENTRADA: string[10];
ETA: array [1..10] of byte;
ANNFILENUM,SEQFILENUM,BT1,BT2,BT3,
FILEINFO,DIVNUM,TYPINFO,ECODE: byte;
RECORDLENGTH,FIRSTLINE,LASTLINE,SEQRECNUM,
MIN,MA,PP,NUM,PRINCIPIO,FIN,DUMMY,D1,D2,
D3,D4,SCORE1,SCORE2,SCO,ANNRECNUM,nbus,
SEGRECLength,ANNRECLength,AA,A,B,C,D,E: integer;
LUG: array [1..200] of integer;
INDEX,SEQUENCE: INDEX_TYP;
BITARY: BITE_TYP;
ADDRESS,BAENUM,FROM,TWO,BB: real;
SALIDA: text;
REP: boolean;
```

```
{*****}
{*          Decodifica secuencia          *}
{*****}
procedure PRINT(BITE:BITE_TYP;var DONE:boolean);
{*          Ver programa LOCALIZA CADENAS          *}
```

```
{*****}
{*          Descompone la información leída en forma compacta para          *}
{*          para su posterior decodificación          *}
{*****}
procedure INFLATE;
{*          Ver programa LOCALIZA CADENAS          *}
```

```

(*****
{*                               Selecciona grupo de lectura
(*****
procedure GET_GROUP;
(*                               Ver programa LOCALIZA CADENAS
*)

(*****
{*                               Calcula apuntadores para lectura
(*****
procedure FIND_POINTER;
(*                               Ver programa LOCALIZA CADENAS
*)

(*****
{*                               Cambia minúscula por mayúscula
(*****
procedure UPER;
(*                               Ver programa LOCALIZA CADENAS
*)

(*****
(*                               Compara secuencias del banco de datos con las secuencias REP
(*                               y escribe el resultado en caso de igualdad
(*****

procedure BUSQUEDA;
begin
REP:=false; CREP1:=0; CREP2:=0; CREP3:=0; CREP4:=0; PP:=0;
for D1:=NUM+1 TO NUM+20 do SEC[D1]:='.';
for D1:=1 to NUM do begin
SCO:=0;
for D2:=0 to 12 do if REP1[D2]=SEC[D1+D2] then SCO:=SCO+1;
if SEC[D1+6]='a' then SCO:=SCO+1;
if SCO>=SCORE1 then begin
PP:=PP+1; LUG[PP]:=D1;
REP:=true; CREP1:=CREP1+1;
write(TEMPFILE,'En 1: ',SCO:2,' de 13');
for D2:=0 to 12 do begin
write(tempfile,sec[d1+d2]);
if REP1[D2]=SEC[D1+D2] then UPER;
end;
writeln(tempfile);
if SEC[D1+6]='a' then SEC[D1+6]:='A';
end;
end;
for D1:=1 to NUM do begin
SCO:=0;
for D2:=0 to 12 do if REP2[D2]=SEC[D1+D2] then SCO:=SCO+1;
if SEC[D1+6]='t' then SCO:=SCO+1;
if SCO>=SCORE1 then begin
PP:=PP+1; LUG[PP]:=D1;

```

```

REP1:=true; GREP2:=CREP2+1;
write(TEMPFILE,'En 2: ',sco:2,' de 13');
for D2:=0 to 12 do begin
  write(tempfile,sec[d1+d2]);
  if REP2[D2]=SEC[D1+D2] then UPER;
end;
writeln(tempfile);
if SEC[D1+6]='t' then SEC[D1+6]:='T';
end;
end;
IF REP THEN BEGIN
for D1:=1 to NUM do begin
  SCO:=0;
  for D2:=0 to 7 do if REP3[D2]=SEC[D1+D2] then SCO:=SCO+1;
  if SEC[D1+3]='t' then SCO:=SCO+1;
  if SCO>=SCORE2 then begin
    cREP3:=cREP3+1;
    for D2:=0 to 7 do begin
      if REP3[D2]=SEC[D1+D2] then UPER;
    end;
    if SEC[D1+3]='t' then SEC[D1+3]:='T';
  end;
end;
for D1:=1 to NUM do begin
  SCO:=0;
  for D2:=0 to 7 do if REP4[D2]=SEC[D1+D2] then SCO:=SCO+1;
  if SEC[D1+4]='a' then SCO:=SCO+1;
  if SCO>=SCORE2 then begin
    cREP4:=cREP4+1;
    for D2:=0 to 7 do begin
      if REP4[D2]=SEC[D1+D2] then UPER;
    end;
    if SEC[D1+4]='a' then SEC[D1+4]:='A';
  end;
end;
end;
d2:=0;
writeln(ENTRYNAME,' ',cREP1,' ',cREP2);
writeln(TEMPFILE,ENTRYNAME,' ',cREP1,' ',cREP2);
for D1:=1 to PP do begin
  write(TEMPFILE,LUG[D1]:6,' ');
  MIN:=LUG[D1]-70; if MIN<1 then MIN:=1;
  MA:=LUG[D1]+70;
  for D2:=MIN to MA do write(TEMPFILE,SEC[D2]);
  writeln(TEMPFILE);
end;
writeln(TEMPFILE);
end;
END;

```

```

(*****
(*   Establece los valores de los vectores REP1 y REP2 de acuerdo *)
(*   al censo de las secuencias REP *)
(*****

procedure ATRIV;
begin
REP1[0]:= 'c'; REP1[1]:= 't'; REP1[2]:= 't'; REP1[3]:= 'a'; REP1[4]:= 't';
REP1[5]:= 'c'; REP1[6]:= 'c'; REP1[7]:= 'g'; REP1[8]:= 'g'; REP1[9]:= 'c';
REP1[10]:= 'c'; REP1[11]:= 't'; REP1[12]:= 'a';
REP2[0]:= 't'; REP2[1]:= 'a'; REP2[2]:= 'g'; REP2[3]:= 'g'; REP2[4]:= 'c';
REP2[5]:= 'c'; REP2[6]:= 'g'; REP2[7]:= 'g'; REP2[8]:= 'a'; REP2[9]:= 't';
REP2[10]:= 'a'; REP2[11]:= 'a'; REP2[12]:= 'g';
REP3[0]:= 'g'; REP3[1]:= 'c'; REP3[2]:= 'c'; REP3[3]:= 'g'; REP3[4]:= 'g';
REP3[5]:= 'a'; REP3[6]:= 't'; REP3[7]:= 'g';
REP4[0]:= 'c'; REP4[1]:= 'a'; REP4[2]:= 't'; REP4[3]:= 'c'; REP4[4]:= 'c';
REP4[5]:= 'g'; REP4[6]:= 'g'; REP4[7]:= 'c';
end;

(*****
(*   Pregunta parametros de corrida *)
(*****
procedure INICIO;
(*   Ver programa LOCALIZA CADENAS *)

(*****
(*   Programa Principal *)
(*   Establece el orden de llamada de las diversas subrutinas *)
(*****

begin
FOUND:=true;
ATRIV;
WRITE('Grupo ? (PRI,ROD,MAM,VRT,INV,PLN,ORG,BCT,RNA,VRL,PHG,SYN,UNA
');
readln(DIVISION);
GET_GROUP_AND_ENTRY_NAME_FROM_USER_AND_OPEN_INDEX_FILE;
assign(TEMPFILE, 'A:TRANS.SEG');
rewrite(TEMPFILE);
writeln(TEMPFILE, DIVISION);
writeln(TEMPFILE);
write('Principio ? ( = ) ', FIRSTLINE, ' ');
readln(PRINCIPIO);
write('Fin ? ( < = ) ', LASTLINE, ' ');
readln(FIN);
write('Score1 (max 13) ');
readln(SCORE1);
write('Score2 (max 8) ');
readln(SCORE2);
for DUMMY:=PRINCIPIO TO FIN do begin
ADDRESS:=((DUMMY-1)/1)*(RECORDLENGTH/1);
longseek(INDEX, ADDRESS);
for A:=1 to 10 do read(INDEX, BTA[A]);

```

```
for A:=1 to 10 do WHY[A]:=chr(BTA[A]);
ENTRYNAME:=WHY;
writeln(DUMMY,' ',ENTRYNAME);
for A:=1 to (10-length(ENTRYNAME)) do SPACE[A]:=' ';
ENTRYNAME:=ENTRYNAME+SPACE;
FIND_SEQUENCE_AND_ANNOTATION_FILES_AND_POSITION_POINTERS;
INFLATE;
BUSQUEDA;
end;
close(TEMPFILE);
end.
```

7.3 Programa 'Elimina redundancia'

```
{*****}
{*          ELIMINA REDUNDANCIA          *}
{*      Dada una secuencia, este programa permite encontrar a      *}
{*      todas aquellas secuencias que, dentro de su vecindad, posean *}
{*      un grado de similitud preestablecido.                    *}
{*****}

program REDUNDAN;

const
    ciclos=30;          (* ciclos *)
    long=16000;         (* longitud de b[] *)
    long2=32000;       (* longitud de a[] *)
    match=20;          (* numero de matches *)

type vec = array [1..long] of byte;
    vec1 = array [1..long2] of byte;

var b: vec;
    a: vec1;
    letras,nombre: string[26];
    palabra: array[1..match] of byte;
    i,j,k,L,m,n,o,p,q,r,s: integer;
    entrada: file of vec;
    salida: text;
    found: boolean;
    chr: array[97..116] of char;

begin
    chr[97]:= 'a'; chr[99]:= 'c'; chr[103]:= 'g'; chr[116]:= 't';
    nombre:= 'frecoli.';
    letras:= 'abcdefghijklmnopqrstuwxxyz';
    assign(entrada, 'coli.seq');
    reset(entrada);
    read(entrada, b);
    for r:=1 to long do a[r+long]:= b[r];
    n:= long div match;
    for i:=1 to ciclos do begin
        if i > 25 then begin nombre[9]:= letras[i-25];
                           nombre[10]:= nombre[9];
                           end
        else nombre[9]:= letras[i];
    assign(salida, nombre);
    rewrite(salida);
    writeln('empieza ciclo ', i, ' ', nombre);
    writeln(salida, i, ' ', nombre);
    read(entrada, b);
    for r:=1 to long do begin a[r]:= a[r+long]; a[r+long]:= b[r]; end;
```



```
writeln;  
for r:=1 to long do a[long+r]:=b[r];  
for j:=1 to n do begin  
  m:=j*match-match;  
  for k:=1 to match do palabra[k]:=a[m+k];  
  q:=m+match; o:=m+long-match;  
  for L:=q to o do begin  
    found:=false; p:=1;  
    repeat  
      if palabra[p]=a[L+p] then p:=p+1 else found:=true;  
    until (found) or (p>match);  
    if not found then begin  
      for s:=1 to match do write(chr[palabra[s]]);  
      writeln;  
      for s:=1 to match do write(salida,chr[palabra[s]]);  
      writeln(salida);  
    end;  
  end;  
  writeln('movimiento: ',j:5,' ( de ',n,' ) del ciclo:  
,1:3, ' ( de  
,ciclos,' )');  
  end;  
  close(salida);  
end;  
end.
```

7.4 Programa 'Busqueda Hash'

```
{*****}
{*          BUSQUEDA.HASH          *}
{*  A partir de una función biunívoca, este programa decodifica  *}
{*  las secuencias de ácidos nucleicos en números. Estos números *}
{*  indican la posición dentro de un vector en donde se calcula  *}
{*  frecuencia de aparición de elementos reiterados.            *}
{*****}
```

program HASH

```
{* Define parámetros de tamaño de acuerdo a la capacidad *}
{* del computador                                        *}
```

```
const MAX=-19873;
      MIN=-32000;
      L=7;
```

type

```
EMPAC=PACKED array [MIN..MAX] of byte;
VECPAC=packed array [0..16383] of byte;
```

```
var INFILE: file of EMPAC;
    SALIDA: file of VECPAC;
    NOMBRES: array [0..15] of string[10];
    FR: array [1..16] of integer;
    NUC: array [0..3] of char;
    O: text;
    BASE: array [1..16] of byte;
    A: EMPAC;
    B: VECPAC;
    CFR,SUPER,A1,A2,L1,J,X1,PO,V,MAX2,P,X,MAXI,
    I,Y,U1,U2,K,Q: integer;
    EX: array [1..7] of integer;
    D: array [0..255,1..4] of byte;
    COUNT: array [0..63] of real;
    SS,S: real;
    RESP: char;
```

procedure DESPLIEGA;

begin

```
Q:=Y div 4096; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*4096;
Q:=Y div 1024; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*1024;
Q:=Y div 256; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*256;
Q:=Y div 64; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*64;
Q:=Y div 16; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*16;
Q:=Y div 4; write(O,NUC[Q]); write(NUC[Q]); Y:=Y-Q*4;
      writeln(O,NUC[Y]); writeln(NUC[Y]);
```

end;

procedure BLANQUEA;

begin

```
NOMBRES[0]:='A'; NOMBRES[1]:='C'; NOMBRES[2]:='G'; NOMBRES[3]:='T';
L1:=L-1;
```

```

PO:=1;
for I:=1 to L-1 do begin
  PO:=PO*4;
  ex[L-I]:=PO;
end;
ex[L]:=1;
MAX2:=MAX-2;
x:=ex[1]*4; x1:=x-1;
for v:=1 to x do B[v]:=0;
for v:=0 to 255 do count[v]:=0;
NUC[0]:='a'; NUC[1]:='c'; NUC[2]:='g'; NUC[3]:='t';
for v:=0 to 63 do d[v,1]:=0;
for v:=64 to 127 do d[v,1]:=1;
for v:=127 to 191 do d[v,1]:=2;
for v:=192 to 255 do d[v,1]:=3;
i:=0;
for v:=0 to 255 do begin
  d[v,4]:=1;
  i:=i+1; if i=4 then i:=0;
end;
i:=0; j:=0;
for v:=0 to 255 do begin
  d[v,3]:=1;
  j:=j+1; if j=4 then begin j:=0; i:=i+1; end;
  if i=4 then i:=0;
end;
i:=0; j:=0;
for v:=0 to 255 do begin
  d[v,2]:=1;
  j:=j+1; if j=16 then begin j:=0; i:=i+1; end;
  if i=4 then i:=0;
end;
end;
end;

{***** programa principal *****}
begin
  blanQuea;
  assign(infile,'lambda.COD');
  reSET(INFILE);
  assign(O,'lambda.frc');
  rewrite(O);
  resp:='s'; u2:=0;
  writeln;
  read(INFILE,A);
  close(infile);
  MAXI:=0;
  a1:=min; a2:=min;

  (* realiza SUPER ciclos de calculo *)
  for SUPER:=0 to 3 do begin
    assign(SALIDA,NOMBRES[SUPER]);
    rewrite(SALIDA);

    (* realiza MAX ciclos de calculo *)

```

```

for v:=min to MAX2 do begin
  k:=0; a2:=a2+1;

  (* decodifica cuatro elementos de tres bytes *)
  for j:=0 to 2 do
    for p:=1 to 4 do begin
      k:=k+1;
      base[k]:=d[a[v+j],p];
    end;

  (* calcula posicion y agrega unidad para cuatro elementos del byte
*)
  for i:=0 to 3 do begin
    if base[i]=SUPER then begin
      p:=0;
      a1:=a1+1;
      for j:=2 to L do p:=p+base[i+j]*ex[J-1];
      B[P]:=B[P]+1;
      if MAXI<B[P] then MAXI:=B[P];
    end;
  end;
  end;
  write(SALIDA,b);
  close(SALIDA);
  writeln('termine con ',NOMBRES[super]);
end;
s:=0; ss:=0;

  (* realiza SUPER ciclos de calculo *)
for SUPER:=0 to 3 do begin
  assign(SALIDA,NOMBRES[SUPER]);
  reset(SALIDA);
  read(SALIDA,b);
  for i:=0 to x1 do begin ss:=ss+B[i]; count[B[i]]:=count[B[i]]+1;
end;
  close(SALIDA);
end;

  (* desplegado de resultado *)
for i:=0 to MAXI do begin
  s:=s+count[i]*i;
  writeln(i:4,' ',count[i]);
  writeln(0,i:4,' ',count[i]);
end;
writeln('frecuencias acumuladas=',s,' ',ss,' a1=',a1,' a2=',a2);
writeln(0,'frecuencias acumuladas=',s,' ',ss,' a1=',a1,'
a2=',a2);
resp:='s'; CFR:=0;
while resp in ['S','s'] do begin
  write('deseas desplegar alguna frecuencia? (S/N) '); read(resp);
  if resp in ['s','S'] then begin
    CFR:=CFR+1;
    write(' Cual frecuencia? '); readln(FR[CFR]);

```

```

end;
end;

if CFR>0 then begin
  (* realiza SUPER ciclos de calculo *)

  for SUPER:=0 to 3 do begin
    assign(SALIDA,NOMBRES[SUPER]);
    reset(SALIDA);
    read(SALIDA,b);
    for j:=1 to CFR do begin
      writeln(O,'frecuencia: ',FR[J]);
      writeln(O,'frecuencia: ',FR[J]);
      for p:=0 to x1 do if L[P]=FR[J] then begin
        write(O,p:7,' ',NUC[super]);
        write(p:7,' ',NUC[super]);
        Y:=p;
        despliega;
      end;
    end;
    close(SALIDA);
  end;
end;
close(O);
END.

```

GENE 1333

Repetitive extragenic palindromic (REP) sequences in the *Escherichia coli* *gdhA* gene

(Regulatory nucleotide sequences; mRNA folding; RNA recognition; nitrogen metabolism; recombinant DNA)

Baltazar Becerril, Fernando Valle, Enrique Merino, Laura Ribá and Francisco Bolívar*

Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Guernavaca, Morelos (México) Tel. (9073)172399

(Received February 1st, 1985)

(Revision received April 1st, 1985)

(Accepted April 11th, 1985)

SUMMARY

Deletions of the 3' flanking DNA region of the glutamate dehydrogenase (GDH) structural gene from *Escherichia coli* K-12, have been produced on a plasmid that carries the complete *gdhA* gene. Those deletions include part of the repetitive extragenic palindromic (REP) sequences proposed by Stern et al. [Cell 37 (1984) 1015-1026], as a novel and major feature of the bacterial genome. The effect of these deletions on the final GDH level in the cell, has been determined. A broader compilation, analysis and alternative functions of the REP sequences, is also presented.

INTRODUCTION

Glutamate dehydrogenase (GDH; EC 1.4.1.4) is an important enzyme because it provides a route for incorporation of nitrogen into organic compounds and thus a link between carbohydrate and aa metabolism (Tyler, 1978). We have previously reported the DNA sequence of 1934 bp consisting of 193 bp

upstream of the ATG start codon for the *E. coli* *gdhA* coding sequence, 1341 bp encoding the 447 aa of the GDH monomer and 400 bp downstream of this coding sequence (Valle et al., 1984). We pointed out that an RNA transcript carrying the sequences located at the 3' end of the structural gene (approx. 200 bp), has the potential to form different secondary structures; however, it is not known where the end of the GDH transcription unit is located. One of such secondary structures beginning at position 1352 (8 bp after the TAA codon) of the *gdhA* gene, showed a remarkable homology (28 identities out of 29 bp), with the consensus sequence proposed by Higgins et al. (1982), as a novel regulatory element of prokaryotic genes.

Gilson et al. (1984) have reported a family of dispersed REP sequences in *E. coli*. These se-

* To whom correspondence and reprint requests should be mailed at Apartado Postal 70479, México 04510 D.F. (México).

Abbreviations: aa, amino acid(s); bp, base pairs; EtBr, ethidium bromide; GDH, glutamate dehydrogenase; kb, kilobases or kilobase pairs; ORF, open reading frame; NH medium, non-nitrogen minimal salts medium; nt, nucleotide(s); ^R, resistance; REP sequence, repetitive extragenic palindromic sequence (Stern et al., 1984); Tc, tetracycline.

quences, called palindromic units, have a consensus sequence similar to that of the novel regulatory element of Higgins et al. (1982). Stern et al. (1984), have published a slightly different consensus sequence for this element, which has been designated REP sequence. This extragenic element is about 35 bp long, includes an inverted repeat and can occur singly or in multiple adjacent copies, predominantly with opposite orientations between any pair of them. The orientation, position and distribution of REP sequences do not seem to obey any immediately obvious rule (Stern et al., 1984).

In this paper, we report on the location of two REP sequences at the 3' end of the *gdhA* gene of *E. coli* K-12 that show strong homology with the consensus REP sequence (Stern et al., 1984). We also report on the construction of several plasmids that carry different deletions which include part of the *E. coli* *gdhA* REP sequences and the effect of those deletions on the final GDH level in the cell.

Finally, a broader analysis of the sequences discussed by Stern et al. (1984), alternative possible functions, and the compilation and analysis of new ones, are presented.

MATERIALS AND METHODS

(a) Bacterial strains and plasmids

E. coli strains HB101 [F^- , *hsdS20* ($r_{II}^- m_{II}^-$), *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20* (Sm^R), *xyl-5*, *mtl-1*, *supE44*, λ^-] and the glutamate auxotroph PA340 (F^- , *leuB6*, *gdh-1*, *htsG1*, *glbB31*, *argH1*, *thi-1*, *ara-14*, *lacY1*, *gal-6*, *malA1*, *xyl-7*, *met-2*, *tonA2*, λ^+ , λ^- , *rpsL9*, *supE44*), were used to transform plasmid constructions as described by Cohen et al. (1972). Plasmid pSAE4 has been previously described (Sanchez-Pescador et al., 1982; Valle et al., 1983).

(b) Enzymes

StuI and *NruI* restriction endonucleases were purchased from New England Biolabs and used in conditions recommended by the suppliers. T4 DNA ligase was purified and used according to the method described by Tait et al. (1980).

(c) Plasmid constructions and copy number estimations

Plasmid pSAE4, which carries the *gdhA* gene and flanking DNA regions, was digested with the restriction endonucleases *StuI* or *StuI* + *NruI*, as to delete 1000 bp or 3000 bp, respectively, from the 3' end of the gene. The resulting plasmids were initially characterized by restriction mapping of plasmid DNA obtained from HB101 strain by the rapid procedure of Rodriguez and Tait (1983). This DNA was used to transform the PA340 strain.

Plasmid copy number was estimated by extraction of plasmid DNA, cleaving at a single site, and separation on an agarose gel. After electrophoresis, the gel was stained with EtBr, photographed, and the DNA quantified by densitometric scanning of the negative.

(d) Nucleotide sequences search

The nucleotide sequences search of the GenBank database (September, 1984), was performed using a Burroughs B 7800 computer with several programs implemented in our laboratory.

RESULTS

(a) Structural characteristics in the 3' DNA region of the *gdhA* gene

Some of the palindromic sequences occurring at the 3' end of the *gdhA* gene are presented in Fig. 1. They could form a tRNA-like structure (Fig. 2), with 35% of its nucleotides positioned as in the *E. coli* glutamate tRNA (Sprinzi and Gauss, 1983), and including two opposing REP sequences. These two sequences, starting at positions 1353 and 1420 (Fig. 1), have very strong homology with the consensus REP sequence reported by Stern et al. (1984), and as can be seen in Fig. 1, they can be folded into, at least, two additional modes.

(b) Construction of REP deletion mutants

Stern et al. (1984) have proposed that REP sequences could be involved in the regulation of gene

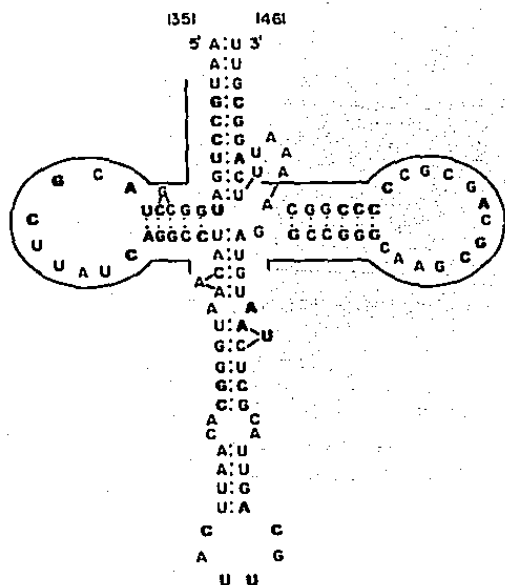


Fig. 2. A trRNA-like structure resulting from folding of the RNA transcript carrying the first 120 nt after TAA stop codon (for numbering see Fig. 1). The nt in common with glutamate tRNA appear in bold characters. The location of REP sequences are indicated by solid lines.

expression in prokaryotes. In an effort to understand the possible role of this kind of sequence in the regulation of GDH concentration in the bacterial cell, we decided to delete one of these REP sequences located at the 3' end of the *gdhA* gene. This experiment was performed taking advantage of the presence of two unique *Sma*I sites located at the 3' region outside the coding region of the *gdhA* gene (positions 1375 and 2375, Fig. 1), and a unique *Nru*I site located in the Tc^R gene of the cloning vehicle pBR322 (Bolivar et al., 1977). Using this strategy, plasmids pBB5 and pBB6 were derived from plasmid pSAE4. These two derivatives were constructed to place different DNA sequences downstream of the *gdhA* gene and so to avoid overinterpretation of the data.

The GDH-specific activities, using *E. coli* strain PA340 as host, were compared for these three plasmids. The final GDH concentrations are reduced to at most 70% (usually 50%) by the deletions, as compared with cells carrying pSAE4 (Fig. 3). It is interesting to point out that this effect is independent of the growth medium. This result is in agreement with the data reported by Stern et al. (1984), who demonstrated that the deletion of the REP sequences in the *hisJ-P* operon of *Salmonella typhimurium*,

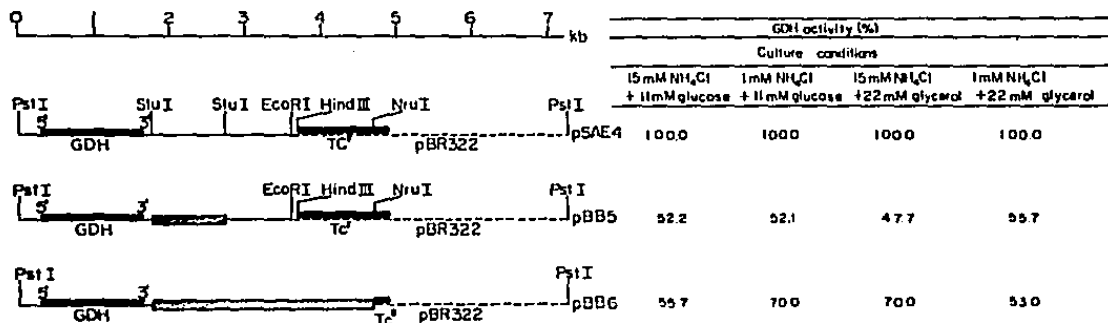


Fig. 3. Physical maps of the plasmids pSAE4, pBB5 and pBB6 linearized at the *Pst*I site. The dashed lines represent the vector pBR322, the solid lines represent *E. coli* chromosomal inserts and the thick bars indicate the locations of the *gdhA* and Tc^R genes. The hatched areas indicate deleted regions. For GDH assays, 5-ml cultures in the supplemented NN media were grown until stationary phase and used to determine the specific activity. The ammonia-dependent activity of GDH was determined as described by Covarrubias et al. (1980). Cells were grown in NN medium supplemented with 1 mM or 15 mM NH₄Cl as the nitrogen source, and 11 mM glucose or 22 mM glycerol, as the carbon source. Absolute values of GDH activity were found to vary slightly from day to day. However, the relative values for any two plasmids assayed on the same day were constant. Therefore, we express all values as percentages rather than in absolute terms. According to our experience with the GDH assay, differences in expression between plasmids of less than $\pm 5\%$ should not be considered to be significant.

TABLE I

Location of REP sequences

The location of each REP sequence, with respect to proximal gene (nearest identified gene to the REP sequences respect to the direction of transcription), is indicated by heavy arrows, whose directions indicate the orientation of such sequences. The orientation of transcription is from left to right. Each gene is represented by a box with its name included. In many cases, it is not known whether additional downstream genes are present. NI (non identified), indicates that an ORF and/or gene product exists but its function is unknown. Those marked with asterisks were found in the present study. S.t. means *Salmonella typhimurium*. In the case of *hisG*, one of the REP sequences is only partially homologous to the consensus (see Higgins et al., 1982). The incomplete *mtlA* REP sequence, indicated by the dashed arrow, may extend further; additional sequence data are not available. There are nine transcriptional units where REP has been located as part of the mRNA: *glyA*, *mulE*, *hisJ*, *mtlA*, *lacY*, *folA*, *hisG*, *glnALG* and *rpoD*. The last two were detected in this study.

Name		Distance from end codon	Separation between REP sequences
<i>oroG</i>		49	—
<i>mtlA</i>		41	—
<i>folA</i>		24	—
<i>lacY</i>		9	—
* <i>dhuA</i> (S.t.)		a	—
* <i>rpoD</i>		3	—
* <i>uvrD</i>		b	—
MI RNA		c	3.5
<i>spot42</i>		60	25
<i>hisJ</i> (S.t.)		47	20
<i>hisG</i> (S.t.)		6	25
<i>glyA</i>		36	3.5
* <i>supBE</i> tRNA		d	5
* <i>tyr</i> tRNA		e	1.3
* <i>gdhA</i>		9	3.4
* <i>glnALG</i>		59	1.4
* <i>plsB</i>		12	3.6
<i>lamB</i>		9	7,14
<i>ndhA</i>		12	1.4
<i>trpR</i>		4	15,26
<i>glnS</i>		20	30,35
<i>sdhA</i>		NA	NA
<i>sucB</i>		NA	NA
* <i>rpoB</i> (S.t.)		14(f)	6,11

* *dhuA* (S.t.), is the name of the regulatory region of the *S. typhimurium* histidine transport operon. In this case, REP sequence overlaps with the promoter proposed for *hisJ*.

^b REP sequence is located between Pribnow box and the initiation ATG codon of the *uvrD* coding region.

^c It has not been established whether a small polypeptide is coded by the region between REP sequence and M1 RNA.

^d REP sequences are located at the 3' end of the tRNA operon. This is one of two cases in which two REP have been found tandemly repeated.

^e REP sequences are located at the 3' end of the tRNA operon.

^f REP sequences located within the intergenic region between *araA* and *araD* genes of the *araBAD* operon of *Salmonella typhimurium* LT1 (Lin et al., 1985). This is another case in which two REP are tandemly repeated.

^{NA} Nucleotide sequences surrounding REP sequences are not available (Stern et al., 1984).

TABLE II

Homology between REP sequences

Name	Sequence
Consensus	T G C C G G A T G C G G C G C 0 to 6 nt G C C T T A T C C G G C C T A C
	T G C C T G A T G C G A C G T A C G T C T T A T C A G G C C T A C
maIE	G C C G G A T G G C G C G T G A A C G C C T T G T C C G G C C T A C
sucB (2)	G C C T G A T G C G A C G C T G A C G C G T C T T A T C A G G C C T A C
sucB (4)	G C C T G A T G T G A C G T T T T A A C G T G T C T T A T C A G G C C T A C
sdhA (2)	G C C G G A T G C G C G T G A G A C G C C T T A T C C G G C C T A C
MI RNA	T G C C G G A T G C G G C G T G A A C G C C T T A T C C G G C C T A C
hisG	C G C C T G A T G G C C G T G C G C C T T A T C C A G G C C T A C
polA	T G C C T G A T G C G C G T A T G T T T A T C A G G C C A A C C
glyA (1)	T G C C G G A T G C G G C G T G A A C G C C T T A T C C G G C C T A C
qdhA	T G C C T G A T G G G C G C A A T A C G C C T T A T C A G G C C T A C
glnALG	C G C C G G A T G C G A C G C A A A T G C G C C T T A T C C G G C C T A C
trpR (2)	T G C C G G A T G C G G C G T G A A G G C C T T A T C C G T C C T A C
hisJ (S.t.)	T G C C T G A T G C G G C G C T T G T G T G T C A G G C C T A C
lamB (2)	C G C C T G A T G C G A C G C T T T G C G C G T C T T A T C A G G C C T A C
glnS (3)	T G C C T G A T G C G G C G T A A A A C G C C T T A T C A G G C C T A C
plsB	T G C C G G A T G C G G C G T A A A A C G C C T T A T C C G G C C T T C
mtIA	T G C C T G A T G C G C G C C G C T T G C G C C T T A T C C G G C C T T C
foiA	C G C C G G A T G C G A C G C C G G T C G C G T C T T A T C C G G C C T T C
lacY	T G T C G G A T G C G G C G C G T G A C G C C T T A T C C G A C C A A C
uvrD	T G G C G G A G G G C G T G C T G C G C T T T C T C C G C C C A A C
tyr tRNA	T G C C C G A T G G C A C C A C G T T T A C A G G G C C T A C
ndhA (2)	T G C C T G A T G C G C T T C T T A T C C G G C C T A C
ndhA (3)	T G G C G G A T G T G G C A T A A A C G C C T T A T C C G C C C T T G
supBE tRNA	T G T C G A A T G C G A T G T T G A C A C G T C T T A T C C
supBE LRNA	T G T C G G A T G C G A C G C T G C C G C G T C T T A T C C G A C C T A C
araBAD (1) (S.t.)	T G C C G G T G G C A C T G C G T T T A C C G G C C T A C

Nucleotide sequences of the consensus REP sequence compared with the REP sequences shown in Table I. The rightward arrow indicates the orientation of REP sequences with respect to the identified proximal gene. The numbers in parentheses identify REP sequences in those units where exist more than two of them. All nucleotide sequence data are from GenBank (September, 1984), except *maIE*, *hisG*, *sucB* and *sdhA*, which are from Stern et al. (1984).

decreased the final concentration of the enzyme coded by the gene located upstream of the REP sequences.

(c) Detailed analysis of the GenBank

These results led us to perform a more detailed study of REP and neighboring sequences. A computer search of the GenBank database (September, 1984), was conducted using the most conserved segment of the REP sequence:

5'-CTTATC^CAGGCCTA-3' and its complemen-
A

tary nucleotide sequence, instead of the sequence used by Stern et al. (1984). These researchers reported a series of REP sequences from which a consensus was derived. Our results are in agreement with theirs, although we would like to emphasize

certain new features. In those transcriptional units where REP has been located, and it is known where transcription ends (9 out of 25), REP is localized within the transcript. Therefore, we decided to compare REP sequences, taking into account their orientation toward the nearest identified gene (proximal gene), in relation to the direction of transcription. By such comparison we were able to add an extra base to the proposed consensus and we also noticed certain features at the central (nonpairing) region of REP sequences. These results are presented in Tables I, II and III. Table I shows the bacterial transcriptional units in which REP sequences have been found. The REP copy number, the orientation with respect to the proximal gene, the distance from the stop codon of this gene, and the distance between REP sequences, are indicated. It is important to notice that some of these REP sequences overlap with proposed promoters, or are located at the end

substrate for specific endonuclease cleavage. They argue that this could be the case because such sequences had not been found in stable RNA species. However, our results show that, at least in *tyr* tRNA and *supBE* tRNA transcriptional units (Table I), REP sequences are present, and also that REP-like sequences were detected in the 16S and 23S and in the *ile* tRNA (Table IV). We believe that the importance of finding REP sequences at the end of tRNA operons, lies in the fact that, at least for *tyr* tRNA and *supBE* tRNA operons, REP sequences could not easily be involved or related to RNA stability. Nevertheless, the finding of this REP and of REP-like sequences in operons that encode stable RNAs does not rule out the possibility that such sequences and/or the sequence 5'-GCAAATTC AATATATTGCA-3', located between two REP sequences, could be involved in the processing of certain mRNA molecules. This hypothesis is supported by the fact that, at least for one case where this sequence has been located (between the 5S RNA and *val* tRNA from *B. subtilis*), it has been proposed that this sequence is involved in the conformation of an RNA processing locus (Gegenheimer and Apirion, 1981).

There are other possible roles for REP sequences. One of them is that they are required to endow specificity for protein binding (Stern et al., 1984). We have already indicated that, at least for *gdhA*, these sequences could be structured as part of a tRNA-like molecule. This might suggest that the molecule could be recognized by a tRNA-modifying-like enzyme. Similar data have been published by Ames et al. (1983), where a tRNA-like structure has been found as part of the leader region of *S. typhimurium his* operon. As in the case of *gdhA*, there is an important homology between *his* tRNA and their tRNA-like molecule. They have also proposed that this tRNA-like structure might have a regulatory function. Furthermore, tRNA-like sequences have been found in several eukaryotic RNA viruses and they are recognized and aminoacylated by host enzymes, despite significant deviation from the consensus tRNA structure (Haenni et al., 1982).

An interesting alternative role for REP sequences as specific protein binding sites comes from the fact that the REP-like sequences that have been found in the 16S and 23S rRNAs (Table IV), have already been involved in the binding of specific ribosomal proteins. The REP-like sequence of the 16S riboso-

mal RNA interacts with the S7 protein (Prince et al., 1983; Noller, 1984); and in the case of 23S ribosomal RNA, the REP-like sequence is part of a structural domain that interacts with L7, L10, L11, L12, and EF-G proteins (Prince et al., 1983). Therefore, an interesting possible role for REP sequences [not included among the several alternative functions proposed by Stern et al. (1984)], could be that these nucleotide sequences do interact with specific proteins, not necessarily nucleases, but with proteins of the general translation machinery of the cell. This would explain why they can be found not only in mRNA molecules, but also in stable 16S, 23S or tRNA molecules. We would like to hypothesize that by the interactions of these REP sequences with specific modifying enzymes and/or specific ribosomal proteins, the rate of translation and/or degradation of certain mRNA molecules could be modulated.

Finally, an alternative role for REP sequences has already been exposed by Stern et al. (1984). They have proposed that REP sequences could be involved in chromosome structure and organization, possibly serving as a binding site for specific proteins. Such proteins may anchor the DNA or may serve to delineate structural or functional domains. However, this proposal could be structured in a different way: it has been demonstrated that REP is part of those RNA transcripts where this sequence has been searched for, thus strongly suggesting that the transcription of this kind of sequence is important.

Worel and Burgi (1972) have shown that the forces which create and stabilize the folded bacterial chromosome are due mainly to RNA-DNA interactions. Therefore, it is possible that the transcription of REP sequences in RNA could contribute to these RNA-DNA interactions.

ACKNOWLEDGEMENTS

We would like to thank Noemi Flores and Francisco Rosseti for their technical support. This work was supported by a grant (PCCBNA - 022584) from the Consejo Nacional de Ciencia y Tecnología, México.

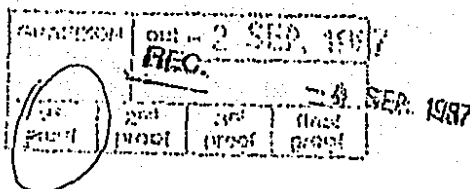
REFERENCES

- Ames, B.N., Tsang, T.H., Buck, M. and Christman, M.F.: The leader mRNA of the histidine attenuator region resembles tRNA^{His}: Possible general regulatory implications. *Proc. Natl. Acad. Sci. USA* 80 (1983) 5240-5242.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Cross, J.H. and Falkow, S.: Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2 (1977) 95-113.
- Cohen, S.N., Chang, A.C.Y. and Hsu, L.: Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *E. coli* by R factor DNA. *Proc. Natl. Acad. Sci. USA* 69 (1972) 2110-2114.
- Covarrubias, A.A., Sanchez-Pescador, R., Osorio, A., Bolivar, F. and Bastarrachea, F.: ColE1 hybrid plasmids containing *Escherichia coli* genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* 3 (1980) 150-164.
- Gegenheimer, P. and Apirion, D.: Processing of prokaryotic ribonucleic acid. *Microbiol. Rev.* 45 (1981) 502-541.
- Gilson, E., Clement, J.M., Brutlag, D. and Hofnung, M.: A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* 3 (1984) 1417-1421.
- Green, C.J. and Vold, B.S.: Sequence analysis of a cluster of twenty-one tRNA genes in *Bacillus subtilis*. *Nucl. Acids Res.* 11 (1983) 5763-5774.
- Haenni, A., Joshi, S. and Chapeville, F., in Cohn, W.E. (Ed.), *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, New York, 1982, pp. 85-102.
- Higgins, C.F., Ames, G.F.L., Barnes, W.M., Clement, J.M. and Hofnung, M.: A novel intergenic regulatory element of prokaryotic operons. *Nature* 298 (1982) 760-762.
- Lin, H.C., Lei, S.P., Studnicka, G. and Wilcox, G.: The *araBAD* operon of *Salmonella typhimurium* LT1. III. Nucleotide sequence of *araD* and its flanking regions, and primary structure of its product, L-ribulose-5-phosphate-4-epimerase. *Gene* 34 (1985) 129-134.
- Noller, H.F.: Structure of ribosomal RNA. *Annu. Rev. Biochem.* 53 (1984) 119-162.
- Prince, J.B., Gutell, R.R. and Garrett, R.A.: A consensus model of the *Escherichia coli* ribosome. *Trends Biochem. Sci.* 8 (1983) 359-363.
- Rodriguez, R.L. and Tait, R.C.: *Recombinant DNA Techniques*. Addison-Wesley, Reading, MA, 1983, pp. 50-51.
- Sanchez-Pescador, R., Sanvicente, E., Valle, F. and Bolivar, F.: Recombinant plasmids carrying the glutamate dehydrogenase structural gene from *Escherichia coli* K-12. *Gene* 17 (1982) 1-8.
- Sprinzi, M. and Gauss, D.H.: Compilation of tRNA sequences. *Nucl. Acids Res.* 12 (1983) r1-r57.
- Stern, M.J., Ames, G.F.L., Smith, N.H., Robinson, E.C. and Higgins, C.F.: Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37 (1984) 1015-1026.
- Tait, R.C., Rodriguez, R.L. and West, R.W.: The rapid purification of T4 DNA ligase from λ T4lg lysogen. *J. Biol. Chem.* 255 (1980) 813-814.
- Tyler, B.: Regulation of the assimilation of nitrogen compounds. *Annu. Rev. Biochem.* 47 (1978) 1127-1162.
- Valle, F., Sanvicente, E., Seeburg, P., Covarrubias, A., Rodriguez, R.L. and Bolivar, F.: Nucleotide sequence of the promoter and amino-terminal coding region of the glutamate dehydrogenase structural gene of *Escherichia coli*. *Gene* 23 (1983) 199-209.
- Valle, F., Becerril, B., Chen, E., Seeburg, P., Heyneker, H. and Bolivar, F.: Complete nucleotide sequence of the glutamate dehydrogenase gene from *Escherichia coli* K-12. *Gene* 27 (1984) 193-199.
- Woreel, A. and Burgi, E.: On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* 71 (1972) 127-147.

Communicated by A.J. Podhajska.

Gene, 00 (1987) 000-000
Elsevier

GEN 02128



Short Communication

Deletion of a repetitive extragenic palindromic (REP) sequence downstream from the structural gene of *Escherichia coli* glutamate dehydrogenase affects the stability of its mRNA

(Regulatory nucleotide sequences; nitrogen metabolism; recombinant DNA)

Enrique Merino, Baltazar Becerril, Fernando Valle and Francisco Bolívar

Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos C.P. 62270 (México)

Received 27 May 1987

Accepted 10 July 1987

URGENT

SUMMARY

A deletion that removes one repetitive extragenic palindromic sequence downstream of the structural gene of *Escherichia coli* glutamate dehydrogenase, reduces twofold the half-life of *gdhA* mRNA.

INTRODUCTION

A novel feature of the *E. coli* and *Salmonella typhimurium* genomes is the presence of numerous copies of the REP sequence (Higgins et al., 1982; Stern et al., 1984; Gilson et al., 1984; Becerril et al., 1985; Newbury et al., 1987). As far as it is known, all REP sequences are in extragenic regions and most of them are located at the 3' terminus of a structural gene (Stern et al., 1984). This 35-bp

sequence has a highly conserved dyad symmetry which could permit the formation of a stem-and-loop structure in its corresponding RNA transcript. This element can occur singly or in juxtaposed copies separated by less than 25 bp; in this case, alternative secondary structures could be formed (Higgins et al., 1982).

The function of REP sequences is still not clear. Since it is present in such a large number of copies per chromosome and is highly conserved in sequence, the element could have an important function for the cell and several possible roles have been proposed. The first one states that REP sequences may be involved in chromosome structure and organization (Stern et al., 1984). The second hypothesis supposes that the REP element could play a role in chromosomal rearrangements and generation of genetic diversity (Higgins et al., 1982; Stern et al., 1984; Gilson et al., 1984). As the third

Correspondence to: F. Bolívar, Apdo. Postal 510-3, Cuernavaca, Morelos C.P. 62270 (México) Tel. (5273) 17-23-99.

Abbreviations: bp, base pair(s); *galK*, galactokinase structural gene; GDH, glutamate dehydrogenase; *gdhA*, GDH structural gene; kb, 1000 bp; nt, nucleotide(s); ^R, resistance; REP, repetitive extragenic palindromic sequence; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl, 0.015 M Na₂ citrate, pH 7.6; Tc, tetracycline; [], designates plasmid-carrier state.

alternative, it has been proposed (Higgins et al., 1982; Stern et al., 1984; Gilson et al., 1984; Plamann and Stauffer, 1985; Becerril et al., 1985) and recently demonstrated (Newbury et al., 1987) that this element may also be involved in the regulation of gene expression. Finally, Newbury et al. (1987) have proposed that REP sequences may be a prokaryotic equivalent of 'selfish' repetitive DNA.

Concerning the third hypothesis, Higgins et al. (1982) and Stern et al. (1984) have suggested that REP sequences could affect gene expression by altering mRNA stability, acting as a barrier to 3' → 5' exonucleolytic degradation of the mRNA (Lafzeron et al., 1976), as a consequence of their secondary (stem-and-loop) structure. Subsequently, it has been reported that the deletion of one (Becerril et al., 1985) or two (Stern et al., 1984) of these downstream elements, or the insertion of different sequences between them (Plamann and Stauffer, 1985), reduces the expression of the upstream gene product. In a recent paper, Newbury et al. (1987) have demonstrated that this effect is mediated at the

level of mRNA stability. The REP element seems to stabilize, in a general fashion, the transcript upstream from it. In a previous paper, our group reported that in *E. coli*, GDH gene expression is reduced almost 50% when one of the two 3' REP sequences of the *gdhA* gene is removed (Becerril et al., 1985). To determine whether this decrease is also mediated by an alteration of the half-life of *gdhA* mRNA, we analyzed the mRNA stability of this transcript.

EXPERIMENTAL AND DISCUSSION

(a) Plasmid constructions and mRNA characterization

The plasmid pBB5 was constructed by deletion of a 1-kb *Sma*I-*Sma*I DNA fragment from plasmid pSAE4 (Fig. 1). As seen in Fig. 2, the *gdhA* mRNA transcripts obtained from both these plasmids are approximately the same size (approx. 1600 nt). The

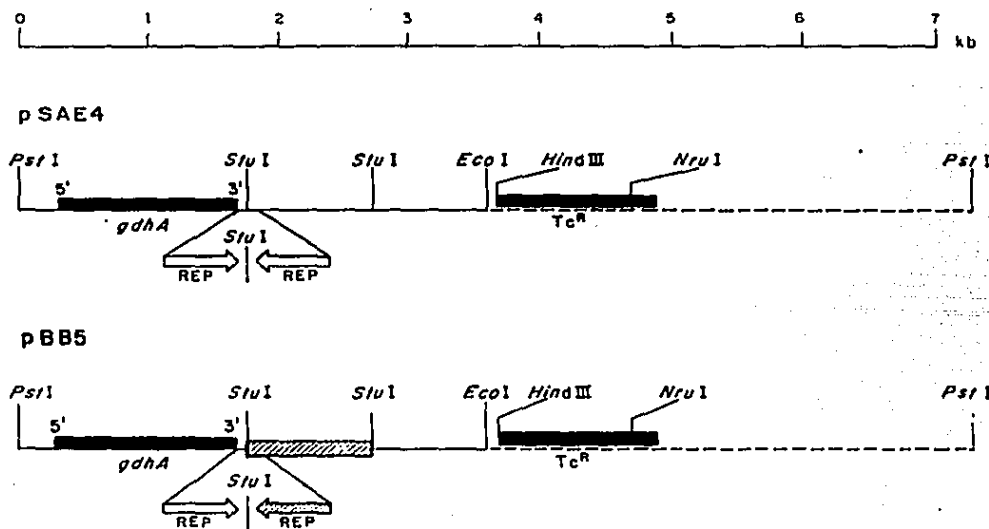


Fig. 1. Physical maps of linearized plasmids pSAE4 and pBB5 carrying the intact *gdhA* gene and flanking regions. Plasmid pBB5 was constructed from plasmid pSAE4 by a deletion of a 1-kb *Sma*I-*Sma*I DNA fragment which includes the second REP sequence (open leftward arrow; Becerril et al., 1985). The dashed lines represent the vector pBR322 (Bolivar et al., 1977; Balbás et al., 1986), the solid lines the *E. coli* chromosomal insert and the thick bars indicate the location of the *gdhA* and *Tc^R* genes. The hatched area indicates the *Stu*I-*Stu*I deletion. The expanded area contains the REP sequences and their orientations, represented by arrows.

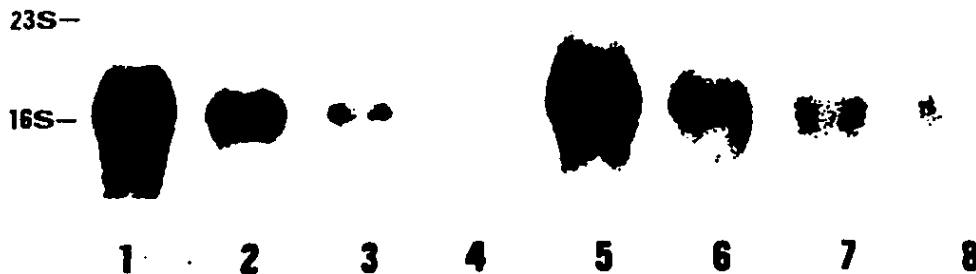


Fig. 2. Determination of the half-life of *gdhA* mRNA. The glutamate auxotroph strain PA340 (Covarrubias et al., 1980), was transformed either with plasmid pSAE4 (Sanchez-Pescador et al., 1982; Valle et al., 1983), which carries the intact *gdhA* gene including the original 3'-flanking region, or with pBB5 (Becerril et al., 1985). In this plasmid the distal REP element was deleted from the sequence flanking the *gdhA* gene (Fig. 1). Transformants were grown to late exponential phase in NN minimal medium (Covarrubias et al., 1980) supplemented with 15 mM NH_4Cl as nitrogen source, and 11 mM glucose as carbon source. After the addition of rifampicin (0.8 mg/ml), total RNA was isolated (Young and Furano, 1981) at intervals (3 min), electrophoresed on a 1% agarose-2.2 M formaldehyde gel for 10 h at 30 V (Lehrach et al., 1977) and then transferred overnight onto nitrocellulose membranes (Southern, 1975). Prehybridization was performed at 42°C for 4 h in a buffer containing 5 × SSC, 50% formamide, 1 mM EDTA, 10 mM Tris pH 8, 10 × Denhardt's solution (Denhardt, 1966) and 150 μg/ml of calf thymus DNA. A fragment of the pBB5 plasmid, carrying the complete *gdhA* gene, was nick-translated according to Rigby et al. (1977) and used as a ^{32}P -labeled probe. Hybridization was carried out for 12 h at 42°C using the prehybridization buffer mixed with the probe. Then the filters were successively washed (twice for 20 min) with: 2 × SSC, 0.1% SDS at room temperature; 0.1 × SSC, 0.1% SDS at 48°C; and 0.1 × SSC at 64°C. The exposure of filters to x-ray films was done at -70°C for 24 h using an intensifying screen (Swanstrom and Shank, 1978). To determine the position of the 23S and 16S rRNA, filters were stained with methylene blue. Lanes 1-4 show *gdhA* transcripts carrying only one REP sequence (from plasmid pBB5) corresponding to 0, 3, 6 and 9 min, respectively, after the addition of rifampicin. Lanes 5-8 show transcripts carrying two REP sequences (from plasmid pSAE4) corresponding to the same times as above. Half-lives were determined by densitometric analysis of the autoradiogram and scintillation counting of the hybridization bands (see Table 1 and Fig. 3).

minimum size for a transcript including the 63 nt of the *gdhA* leader sequence (L. Riba, personal communication) and the 1441 nt of the coding sequence of this enzyme (Valle et al., 1984), is 1504 nt. The similar size of the *gdhA* transcript from both plasmids, suggests that the ends of the transcripts are very near the REP sequence.

(ii) mRNA stability

The mRNA half-life measurements demonstrate that the *gdhA* mRNA carrying the two original REP sequences is about 50% more stable than the *gdhA* mRNA carrying only one REP (Fig. 3). These results are consistent with the levels of GDH activity obtained for each strain (Table I). These data support the hypothesis that REP sequences may affect gene expression by altering mRNA stability. A similar effect on mRNA stability has been previously demonstrated for several operons where two REP sequences were added or removed (Newbury et al.,

1987). The present results demonstrate that removing one of the two REP sequences at the 3' end of a gene is sufficient to reduce the stability of the

TABLE I

GDH specific activity and *gdhA* mRNA half-life

Host(plasmid) ^a	Specific activity of GDH ^b	<i>gdhA</i> mRNA half-life ^c
PA340[pSAE4]	100%	4 min (100%)
PA340[pBB5]	52%	2 min (50%)

^a See Fig. 1.

^b To measure the GDH enzyme activity, 5-ml cultures of strain PA340 containing plasmids pSAE4 or pBB5 were grown to late exponential phase in NN medium supplemented with 15 mM NH_4Cl and 11 mM glucose, and GDH was assayed as described by Covarrubias et al. (1980). Absolute values of GDH-specific activity were found to vary slightly from day to day. However, the relative values for the two plasmids assayed in the same experiment were constant. Therefore, the values are presented in % rather than in absolute terms.

^c See legends to Figs. 2 and 3.

in their

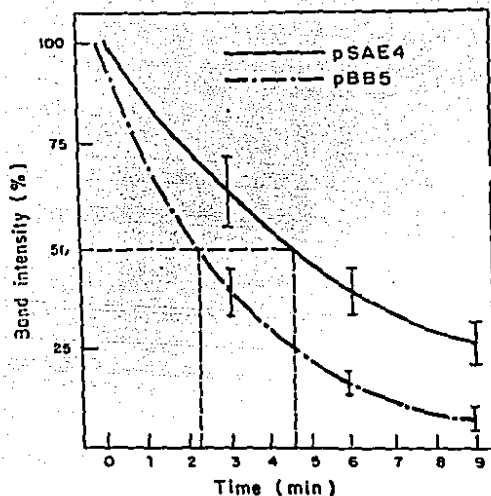


Fig. 3. Degradation kinetics of *gdhA* mRNA. mRNA half-lives were calculated from least-squares-fitted plots of the logarithmic value of the bands' intensity against time, by both densitometric scanning of the x-ray film and scintillation counting of the hybridization bands shown in Fig. 2. The amounts of hybridized probes are expressed in relative terms. Half-lives were graphically calculated (dashed lines), and approximately correspond to $t_{1/2} = 4$ min for the PA340[pSAE4] *gdhA* mRNA and $t_{1/2} = 2$ min for the PA340[pBB5] *gdhA* transcript. The data shown in this figure have been obtained from the averaged values of three independent experiments. Standard deviations are represented by vertical lines.

corresponding transcript. Stern et al. (1984) have demonstrated that insertion of a single copy of the REP sequence causes a reduction in galactokinase activity when it is upstream from the *galK* gene. We believe that this effect is not necessarily related to the mRNA stabilization effect observed when REP sequences are placed downstream from a gene.

It is known that hairpin structures located either at the 3' or 5' end of various genes can stabilize their RNA transcripts (Lozeron et al., 1976; Guarneros et al., 1982; Schmeissner et al., 1984; Hayashi and Hayashi, 1985; Mott et al., 1985; Wong and Chang, 1986; Belasco et al., 1985; Gorski et al., 1985). However, none of these sequences seem to possess such a large or highly conserved palindromic sequence as the REP element. It has also been reported that mutagenesis on the flanking regions of certain

genes may cause a modification in mRNA stability (George and Melton, 1986). With all these evidences, we tend to agree with Stern et al. (1984), Plamann and Stauffer (1985) and Newbury et al. (1987) in the proposition that the REP effect stability on mRNA may not be the primary function of this element. However, in certain situations this property of REP sequences could have been selected by the organism to serve a physiological function.

ACKNOWLEDGEMENTS

This work was partially supported by grant PCCBNAL-022584 from The Consejo Nacional de Ciencia y Tecnologia (CONACyT), México. E. Merino was supported by a fellowship from the CONACyT, México.

REFERENCES

- Balás, P., Sobéron, X., Merino, E., Zurita, M., Lomell, H., Valle, F., Flores, N. and Bolívar, F.: Plasmid vector pBR322 and its special-purpose derivatives — a review. *Gene* 50 (1986) 3-40.
- Becerril, B., Valle, F., Merino, E., Riba, L. and Bolívar, F.: Repetitive extragenic palindromic (REP) sequences in the *Escherichia coli gdhA* gene. *Gene* 37 (1985) 53-62.
- Belasco, J.G., Beatty, J.T., Adams, C.W., von Gabain, A. and Cohen, S.N.: Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polyclonally transcribed *rxcA* transcript. *Cell* 40 (1985) 171-181.
- Belasco, J.G., Nilsson, G., von Gabain, A. and Cohen, S.N.: The stability of *E. coli* gene transcripts is dependent on determinants localized to specific mRNA segments. *Cell* 46 (1986) 245-251.
- Bolívar, F., Rodríguez, R.L., Greene, P.J., Bettlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S.: Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2 (1977) 95-113.
- Blundell, M. and Kennell, D.: Evidence for endonucleolytic attack in decay of *lac* messenger RNA in *Escherichia coli*. *J. Mol. Biol.* 83 (1974) 143-161.
- Covarrubias, A.A., Sanchez-Pescador, R., Osorio, A., Bolívar, F. and Bastarrachea, F.: ColE1 hybrid plasmids containing *Escherichia coli* genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* 3 (1980) 150-164.
- Denhardt, D.T.: A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* 23 (1966) 641-646.

- George, S.E. and Melton, T.: Cloning and molecular characterization of *cim* mutations allowing expression of catabolite-repressible operons in the absence of exogenous cyclic AMP. *J. Bacteriol.* 166 (1986) 533-540.
- Gilson, E., Clement, J.-M., Brullag, D. and Hofnung, M.: A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* 3 (1984) 1417-1421.
- Gorski, K., Roch, J.-M., Prentki, P. and Krisch, H.M.: The stability of bacteriophage T4 gene 32 mRNA: a 5' leader sequence that can stabilize mRNA transcripts. *Cell* 43 (1985) 461-469.
- Guarneros, G., Montañez, C., Hernandez, T. and Court, D.: Posttranscriptional control of bacteriophage λ *int* gene expression from a site distal to the gene. *Proc. Natl. Acad. Sci. USA* 79 (1982) 238-242.
- Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clement, J.-M. and Hofnung, M.: A novel interclastronic regulatory element of prokaryotic operons. *Nature* 298 (1982) 760-762.
- Hayashi, M.N. and Hayashi, M.: Cloned DNA sequences that determine mRNA stability of bacteriophage ϕ X174 in vivo are functional. *Nucl. Acids Res.* 13 (1985) 5937-5948.
- Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H.: RNA molecular weight determinations by gel electrophoresis under denaturing conditions: a critical reexamination. *Biochemistry* 16 (1977) 4743-4751.
- Lozeron, H.A., Dahlberg, J.E. and Szybalski, W.: Processing of the major leftward mRNA of coliphage lambda. *Virology* 71 (1976) 262-277.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Mott, J.E., Galloway, J.L. and Platt, T.: Maturation of *Escherichia coli* tryptophan operon mRNA: evidence for 3' exonucleolytic processing after rho-dependent termination. *EMBO J.* 4 (1985) 1887-1891.
- Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F.: Stabilization of translationally active mRNA by prokaryotic REP sequences. *Cell* 48 (1987) 297-310.
- Plamann, M.D. and Stauffer, G.V.: Characterization of a *ci*-acting regulatory mutation that maps at the distal end of the *Escherichia coli* *gla* gene. *J. Bacteriol.* 161 (1985) 650-654.
- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P.: Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113 (1977) 237-251.
- Sanchez-Pescador, R., Sanvicente, E., Valle, F. and Bolívar, F.: Recombinant plasmids carrying the glutamate dehydrogenase structural gene from *Escherichia coli* K-12. *Gene* 17 (1982) 1-8.
- Schmelssner, U., McKenney, K., Rosenberg, M. and Court, D.: Removal of a terminator structure by RNA processing regulates *int* gene expression. *J. Mol. Biol.* 176 (1984) 39-53.
- Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98 (1975) 503-517.
- Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F.: Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37 (1984) 1015-1026.
- Swanstrom, R. and Shank, P.R.: X-ray intensifying screens greatly enhance the detection by autoradiography of the radioactive isotopes 32 P and 125 I. *Anal. Biochem.* 86 (1978) 184-192.
- Valle, F., Becerril, B., Chen, E., Seeburg, P., Heyneker, H. and Bolívar, F.: Complete nucleotide sequence of the glutamate dehydrogenase gene from *Escherichia coli* K-12. *Gene* 27 (1984) 193-199.
- Wong, H.C. and Chang, S.: Identification of a positive retro-regulator that stabilizes mRNAs in bacteria. *Proc. Natl. Acad. Sci. USA* 83 (1986) 3233-3237.
- Young, F.S. and Furano, A.V.: Regulation of the synthesis of *E. coli* elongation factor Tu. *Cell* 24 (1981) 695-706.

Communicated by Z. Hradečná.

A compilation of Repetitive Palindromic (REP) Sequences

Enrique Merino and Francisco Bolívar*

**Centro de Investigación sobre Ingeniería Genética y Biotecnología.
Universidad Nacional Autónoma de México. Cuernavaca, Morelos, México**

* To whom correspondence and reprint requests should be addressed at
Apdo. Postal 510-3, Cuernavaca, Morelos C.P. 62270 (México).

A novel feature of the Escherichia coli and Salmonella typhimurium genomes is the presence of a highly conserved extragenic palindromic (REP) sequence, found in more than 25% of the operons where it has been searched for. The main features of the REP sequences have been previously described (1, 2, 3, 4, 5, 6). In order to perform an updated compilation of the REP sequences, we searched for the most conserved segment of this element (5'-CTTATCAGGCCTAC-3') and its complementary nucleotide sequence on the GenBank data base (release 44.0). We found 21 new REP sequences in 11 different operons. These data increase the number of reported REP sequences to 83.

ACKNOWLEDGMENTS

We wish to thank the Instituto de Investigaciones Eléctricas, Cuernavaca, Morelos, México, for allowing us to use their computer facilities. We are grateful to Paulina Balbás and Baltazar Becerril for valuable discussions. This work was partially supported by grant FCCBNAL-022584 from Consejo Nacional de Ciencia y Tecnología (CONACyT) México. E. Merino was supported by a fellowship from CONACyT, México.

REFERENCES

1. Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clément, J.M. and Hofnung, M. (1982) Nature 298, 760-762.
2. Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) Cell 37, 1015-1026.
3. Gilson, E., Clément, J.-M., Brutlag, D. and Hofnung, M. (1984) EMBO J. 3, 1417-1421.
4. Becerril, B., Valle, F., Merino, E., Riba, L. and Bolívar, F. (1985) Gene 37, 53-62.
5. Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F. (1987) Cell 48, 297-310.
6. Merino, E., Becerril, B., Valle, F. and Bolívar, F. (1987) Gene In press.
7. Valentin-Hansen, P., Hammer-Jespersen, K., Boetius, F., and Svensen, I. (1984) EMBO J. 3, 179-183.
8. Ueno-Nishio, S., Mango, S., Reitzer, L.J. and Magasanik, B. (1984) J. Bacteriol. 160, 379-384.
9. MacFarlane, S.A. and Merrick, M. (1985) Nucl. Acids Res. 13, 7591-7606.
10. Plamann, M.D., Stauffer, L.T., Urbanowski, M.L. and Stauffer, G.V. (1983) Nucl. Acids Res. 11, 2065-2075.

11. Plamann, M.D. and Stauffer, G.V. (1985) *J. Bacteriol.* 161, 650-654.
12. Clément, J.M. and Hofnung, M. (1981) *Cell* 27, 507-514.
13. Yazyu, H., Shiota-Niia, S., Shimamoto, T., Kanasawa, H., Futai, M., and Tsuchiya, T. (1984) *J. Biol. Chem.* 259, 4320-4326.
14. Saint-Girons, I., Duchange, N., Cohen, G.N. and Zakin, M. (1984) *J. Biol. Chem.* 259, 14282-14285.
15. Duchange, N., Zankin, M.M., Ferrara, P., Saint-Girons, I., Park, I., Tran, S.V., Py, M.-C. and Cohen, G.N. (1983) *J. Biol. Chem.* 258, 14868-14871.
16. Lee, C.A. and Saier, M.H. Jr. (1983) *J. Biol. Chem.* 258, 10761-10767.
17. Markaroff, C.A. and Zalkin, H. (1985) *J. Biol. Chem.* 260, 10378-10387.
18. Erickson, B.D., Burton, Z.F., Watanable, K.K. and Burges, R.R. (1985) *Gene* 40, 67-78.
19. Darlison, M.G., Spencer, M.E. and Guest, J.R. (1984) *Eur. J. Biochem.* 141, 351-359.
20. Spencer, M.E., Darlison, M.G., Stephens, P.E., Duckenfield, I.K., and Guest, R. (1984) *Eur. J. Biochem.* 141, 361-374.
21. Parsot, C., Cossart, P., Saint-Girons, I. and Cohen, G.N. (1983) *Nucl. Acids Res.* 11, 7331-7345.
22. Buvinger, E.W., Lampel, K.A., Bojanowski, R.J. and Riley, M. (1984) *J. Bacteriol.* 159, 618-623.
23. Urbanowski, M.L., and Stauffer, G.V. (1985) *Nucl. Acids Res.* 13, 673-685.
24. Broome-Smith, J.K., Edelman, A., Yousif, S. and Spratt, B.G. (1985) *Eur. J. biochem* 147, 437-446.

Gene	Reference ^a				Distance from stop codon ^b	Separation between REP sequences ^c				
alkA		alkA	→	←	7	14				
araA(St)	4	araB	→	←	←	araD	14	6,11		
aroG	2,3,4	aroG	←				49	-		
deoC	3,7	deoC	→	←	deoA		13	15		
dhxA(St)	4	argI	←	←	hisJ			a		
fnlA	2,3,4	folA	→				23	-		
gdhA	3,4,6	gdhA	→	←			8	b		
glnA	4,5,8,9	glnA	→	←	glnL	glnG	68	c		
glnS	2,3,4	glnS	←	←	←		20			
gln tRNA	3,4	gln tRNA	→	→			51	d		
glyA	2,3,4,10,11	glyA	→	←			38	e		
hisJ(St)	1,2,3,4,5	hisJ	→	←	hisB	hisM	hisP	47	f	
hisG(St)	1,2,3,4	hisG	←	←	hisD	hisC	hisB	6	g	
infB		infB	←	→			13			
lacA		lacA	→				11			
lacY		?	→	←				h		
lacY	2,3,4	lacZ	→	←	lacA		8			
lamB	1,2,3,4,12	malK	←	→	←	molA	9			
lsp		ileS	→	←			7			
malE	2,3,4	malE	→	←	→	malF	malG	10		
malB	13	malB	→				22			
malJ	3,5,14	malJ	→	←			30	i		
malL	3,15	malL	←	←	malF		59			
mtIA	2,3,4,16	mtIA	→	←	mtIB		40			
MIRNA	2,3,4	MIRNA	→	←				j		
ndhA	2,3,4		←	→	ndhA		-			
ndhA	2,3,4	ndhA	→				11			
nrDA		nrDA	→	←	→	←	→	nrDB	11	10,5,11,0
nrDB		nrDB	→				264			
phr		phr	→				27			
plsB	3,4	plsB	→	←			11			
ponA		ponA	→				53	k		
pstA		pstA	→	←	pstB		97			
purF	17		←	←	?	purF		l		
rpoD	3,4,5,10	rpoD	←				3	m		
sdhB	2,4,19	sdhA	←	→	←		3			
spot42	2,3	polA	→	→	→		24	n		
sucB	2,4,20	sucA	←	←	←	→	19			
thrC	3,21	thrB	→	←			543	o		
trpR	1,2,3,4	trpR	←	→	←		4			
tyr tRNA	4	tyr tRNA	→	←				p		
uvrD	4		→	→	uvrD			q		
uvrD		uvrD	→	←			3			

Figure 1. Location of REP sequences. The position of each REP sequence with respect to the upstream proximal gene is represented by an arrow (not to scale) which direction indicates the orientation of such sequence. All the sequences are from Escherichia coli, except those from Salmonella typhimurium that marked with (St).

- a This REP sequence overlaps the proposed promoter for hisJ.
- b A deletion that removes the second REP sequence downstream of the structural gene, reduces twofold the GDH-specific activity (4), due to a decrease of the gdhA mRNA half-life (6).
- c The intercistronic glnA-glnL regions from Escherichia coli, Salmonella typhimurium and Klebsiella pneumoniae are highly homologous among them, nevertheless, only the intercistronic region from Escherichia coli carries REP sequences (5,8,9).
- d The first REP sequence is located at 51 bp from the end of the gln tRNA operon.
- e An insertion of Mu cts phage between the glyA translation stop codon and the REP sequences, reduces the serine hydroxymethyltransferase activity to 30% of the normal value (11).
- f A deletion that removes the two REP sequences from the intergenic hisJ-hisQ region, reduces the level of the hisJ product by about 50% (2) due to a destabilization effect at the mRNA level (5).
- g Only the first half of the second REP sequence has been reported.
- h Escherichia coli DNA region localized 3.3 kb downstream of the lacA gene (22).
- i Despite sequence conservation between the metJ genes from Escherichia coli and Salmonella typhimurium, there are not REP sequences downstream the metJ gene from S. typhimurium (5,14,23).
- j It has not been established whether a small polypeptide is encoded by the region between REP sequences and MIRNA.
- k The reported sequence ends at the second half of the REP sequence (24).
- l Deletion of the REP sequences upstream the structural gene, plays no role in the expression of purF (17).
- m Despite sequence conservation between the rpoD genes from Escherichia coli and Salmonella typhimurium, there is only one REP sequence downstream the rpoD gene from the first organism (5,18).
- n In this case, the distance was considered from the end of the spot42. There is an unidentified gene reading towards spot42, localized 20 pb from REP sequences.
- o There is an open reading frame (ORF127) in opposite orientation of the threonine operon. The stop codon of this putative gene is located 12 pb from the first REP sequence.
- p REP sequences localized at the 3' end of the tyr tRNA operon.
- q The first REP sequence is located between the Pribnow box and the initiation ATG codon of the uvrD coding region.
- ♠ This column indicates the references where a palindromic sequence has been reported as REP element or shares homology with it. All the nucleotide sequences are from GenBank data base (release 44.0) except hisG, sucB, and sdhA (2) and metJ (14).
- * Separation distances are indicated in base pairs.
- * REP sequences found in this study.
- + REP sequences that have been described but not previously compiled.
- ? The function of the gene product has not been identified.

The ribonucleoside diphosphate reductase (nrdA) gene of Escherichia coli carries a Repetitive Extragenic Palindromic (REP) sequence in its 3' structural terminus.

Enrique Merino and Francisco Bolivar*

Centro de Investigación sobre Ingeniería Genética y Biotecnología
Universidad Nacional Autónoma de México. Cuernavaca, Morelos, México

*To whom correspondence and reprint requests should be addressed at Apdo.
Postal 510-3, Cuernavaca, Morelos C.P. 62270 (México).

An additional interesting feature is that usually not conserved inter REP DNA sequences, as well as the "loop" forming regions of these REP sequences, are highly conserved in this case. Newbury et al. (6) have proposed that REP sequences became spread throughout the genome by a "selfish" process, possibly RNA-mediated gene conversion. However, the data presented in this paper suggest that at least for the nrda gene this is not the case, where a possible duplication of its 3' terminus region, apparently generated two REP sequences.

ACKNOWLEDGMENTS

We thank the Instituto de Investigaciones Eléctricas, Cuernavaca, Morelos, México, for allowing us to use their computer facilities. We are also grateful to Paulina Balbás and Baltazar Becerril for valuable discussions. This work was partially supported by grant PCCBNAL-022584 from Consejo Nacional de Ciencia y Tecnología (CONACyT) México. E. Merino was supported by a fellowship from CONACyT, México.

REFERENCES

1. Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clément, J.M. and Hofnung, M. (1982) *Nature* 298, 760-762.
2. Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) *Cell* 37, 1015-1026.
3. Gilson, E., Clément, J.-M., Brutlag, D. and Hofnung, M. (1984) *EMBO J.* 3, 1417-1421.
4. Becerril, B., Valle, F., Merino, E., Riba, L. and Bolivar, F. (1985) *Gene* 37, 53-62.
5. Gilson, E., Perrin, D., Clément, J.M., Szmalcman, S., Dassa, E. and Hofnung, M. (1986) *FEBS Lett.* 206, 323-328.
6. Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F. (1987) *Cell* 48, 297-310.
7. Merino, E., Becerril, B., Valle, F. and Bolivar, F. (1987) *Gene* In press.
8. Gilson, E., Clément, J.M., Perrin, D. and Hofnung, M. (1987) *Trends Genet.* 3, 226-230.
9. Gilson, E., Perrin, P., Saurin, W. and Hofnung, M. (1987) *J. Mol. Evol.* 25, 371-373.
10. Carlson, J., Fuchs, J.A., and Messing, J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4294-4297.