

00581

1 10/1

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

CARACTERIZACION DE LAS REGIONES DE REGULACION DEL GENE ESTRUCTURAL
DE LA ENZIMA GLUTAMATO DESHIDROGENASA DE ESCHERICHIA COLI K-12.

T E S I S

QUE PARA OBTENER EL TÍTULO DE :

DOCTOR EN CIENCIAS QUÍMICAS (BIOQUÍMICA)

P R E S E N T A:

B A L T A Z A R B E C E R R I L L U J A N

México, D.F.

1986

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 esta 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.

I N D I C E

	Página
Jurado Designado	A
Glosario y abreviaturas	E
Resumen en Español	G
" " Inglés	H
INTRODUCCION	1
OBJETIVOS	13
MATERIALES Y METODOS	13
Purificación de ARN total	14
Marcaje de oligonucleótidos con ³² P	16
Extensión de "primer"	17
Mutagénesis de la región regulatoria	19
RESULTADOS	24
Región regulatoria 5' (Mapeo del promotor)	25
Región estructural. Secuencia nucleotídica y de aminoácidos	31
Frecuencia de uso de codones	36
Comparación entre las secuencias de aminoácidos de diversas GDH	36
Región flanqueadora 3'. Características estructurales	37
Construcción de mutantes por eliminación de las secuencias REP	41
Análisis detallado del banco de secuencias (GenBank)	43
DISCUSION	49
PERSPECTIVAS	64
REFERENCIAS	66
PUBLICACIONES	74

ntnC = glnG = Gene que codifica para una protefna con una funci3n regulatoria dual, es decir, como represor o activador, directa o indirectamente.

pb = pares de bases.

pmole = picomole (10^{-12} mol)

^{32}P = Is3toto radiactivo del f3sforo.

Pi = Fosfato inorg3nico libre (HPO_4^{2-}).

rpm = Revoluciones por minuto.

Subunidad sigma (σ) = Subunidad regulatoria de la ARN polimerasa.

SDS = Dodecil-sulfato-de sodio.

Tc^{R} = Resistencia a la Tetraciclina.

V/V = Volumen a volumen.

U = Unidad de actividad especifica.

V = Volt o Voltio.

Vols. = Vol3menes.

' = "prima".

' = minutos.

h = Horas.

nm = nanometros (10^{-9} m).

Ci = Curie o Curio.

CH_3HgOH = Hidr3xido de metil-mercurio.

mA = miliamperes.

NaOAc = Acetato de sodio.

NADPH = Forma reducida del Nicotinamida-adenin-dinucle3tido-fosfato.

NADP^+ = " oxidada " " " " " " "

NADH = Forma reducida del Nicotinamida-adenin-dinucleotido.

NAD^+ = " oxidada " " " " " " "

NN = Medio m3nimo de sales no suplementado.

RESUMEN

Mediante las técnicas descritas por Heidecker et al., 1980 y Messing et al., 1981, se determinó la secuencia nucleotídica de un fragmento de ADN cromosomal de Escherichia coli K-12, el cual porta el gene que codifica para la enzima glutamato deshidrogenasa (GDH) y sus regiones flanqueadoras 5' y 3'.

Un análisis de la secuencia correspondiente a los 200 pares de bases anteriores al codón de iniciación permitió detectar la presencia de varios posibles promotores.

Por medio de la técnica de extensión de primero, se demostró que sólo uno de ellos es funcional en las condiciones metabólicas ensayadas, tanto en plásmido como en cromosoma. La utilización de esta técnica para el mapeo del promotor en plásmidos a los que se les modificó la región regulatoria 5' del gene de la GDH, usando diferentes enzimas de restricción, permitió concluir que sólo un promotor es funcional.

El análisis de la región comprendida en los 200 pares de bases posteriores al codón de terminación, reveló que un transcrito que porte esta región, pudiera tener la capacidad de formar diversas estructuras secundarias. Entre ellas, se encuentra un par de secuencias denominadas en la literatura secuencias extragénicas palindrómicas repetitivas (secuencias REP), propuestas como un aspecto novedoso y relevante del genoma bacteriano.

Se produjeron diferentes deleciones de la región flanqueadora 3' del gene de la GDH en un plásmido que porta el gene completo. Estas deleciones incluyen parte de las secuencias REP. Se determinó el efecto de dichas deleciones sobre la concentración final de la GDH intracelular.

También se presenta una compilación de secuencias REP que amplía las reportadas, así como un análisis y posibles funciones de esas secuencias.

SUMMARY

A chromosomal DNA segment carrying the coding and flanking regions of the glutamate dehydrogenase (GDH) gene from Escherichia coli K-12 has been sequenced

The existence of various possible promoters has been revealed by an analysis of the 200 bp before the initiation codon.

It has been demonstrated by primer extension that only one of - those possible promoters is functional at least in the assayed metabolic conditions, either in plasmid or chromosome.

The analysis of the region comprised of 200 bp after stop codon has permitted to propose that an RNA transcript carrying these sequences, has the potential to form different secondary structures. Between them, a pair of sequences denominated in the literature as repetitive extragenic palindromic (REP) sequences and proposed as a novel and mayor feature of the bacterial genome, was found.

Deletions of the 3' flanking DNA region of the glutamate dehydrogenase structural gene have been produced on a plasmid that - carries the complete gene. The effect of these deletions on the - final GDH intracelular level, has been determined. A broader compilation, analysis and alternative functions of the REP sequences, is also presented.

INTRODUCCION.

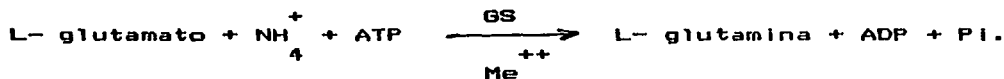
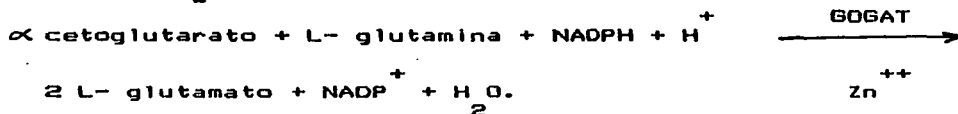
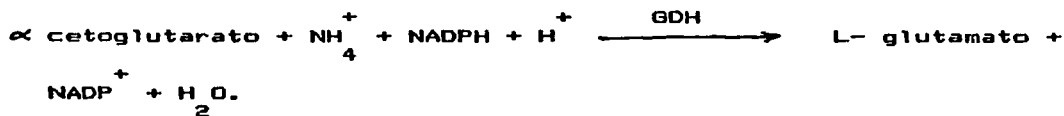
El nitrógeno es uno de los elementos que han sido seleccionados por la naturaleza para formar parte de las moléculas orgánicas (Wald, G., 1962). Siendo éste un elemento esencial para la síntesis de las moléculas nitrogenadas, ha sido de fundamental importancia que los sistemas biológicos hayan desarrollado mecanismos eficientes de asimilación y distribución del nitrógeno del medio circundante.

Los microorganismos utilizan al ión amonio (NH_4^+) y una gran variedad de compuestos nitrogenados como fuentes de nitrógeno. El rango de posibilidades de utilización abarca desde compuestos inorgánicos como el nitrógeno atmosférico (N_2) y el nitrato, hasta compuestos orgánicos complejos tales como aminoácidos (histidina, prolina, arginina) y otros. Todos estos compuestos pueden ser convertidos en amonio por medio de enzimas específicas.

Los mecanismos de asimilación y distribución de nitrógeno han sido estudiados en los microorganismos con cierto detalle. Se sabe que estos crecen y se mantienen en un medio mínimo conteniendo glucosa y amonio como únicas fuentes de carbono y nitrógeno respectivamente (Tyler, B., 1978).

En estas condiciones de crecimiento, los productos primarios de la asimilación del nitrógeno, son el glutamato y la glutamina. Estos últimos ocupan posiciones centrales en el metabolismo intermediario (Stewart et al., 1980).

En Escherichia coli, el glutamato es sintetizado predominantemente por dos enzimas, la glutamato deshidrogenasa (GDH, L- glutamato: NADP óxido reductasa, EC 1.4.1.4.) y la glutamato sintasa (GOGAT, L- glutamato: NADP óxido reductasa, EC (2.6.1.15), mientras que la glutamina es sintetizada por la glutamino sintetasa (GS, L- glutamato amonio ligasa, EC 6.3.1.2.). Las reacciones catalizadas por estas enzimas son respectivamente:



La glutamato deshidrogenasa (GDH), juega un papel central en el metabolismo, en virtud de que es la responsable de la interconversión de α cetoglutarato y glutamato, situandose esta reacción en un punto de cruce entre el metabolismo nitrogenado y el de carbohidratos. Por un lado, el glutamato es el donador primario de grupos amino para la síntesis de aminoácidos por transaminación, y por otro, el α cetoglutarato entra al ciclo de Krebs, generando equivalentes reductores (Prussiner et al., 1972).

Experimentos hechos en K. aerogenes y K. pneumoniae, indican que existe una correlación entre los niveles de GS y la síntesis de otras enzimas involucradas en el metabolismo nitrogenado

(Brenchley et al., 1973; Prival et al., 1973; Streitcher et al., 1976; Tyler, B., 1978; Clarke, P.H., 1979). Cuando K. aerogenes es crecida en un medio mínimo conteniendo glucosa y un exceso de amonio, el glutamato es sintetizado a través de la GDH; y bajo estas condiciones, las actividades de la GS y la GOGAT son bajas (Brenchley et al., 1973; Prival et al., 1973). Si el suministro de amonio es bajo, la GDH es inactiva debido a que posee una K_M alta para este último, siendo entonces incorporado por el par GS-GOGAT, ya que en el caso de la GS, ésta posee una K_M baja para el amonio. En estas condiciones de bajo amonio, los niveles de GS son altos.

Por otro lado, cuando el gene estructural de la GS (glnA) de E. coli, es introducido en el citoplasma de una cepa de K. aerogenes con una mutación en glnA, da como resultado la síntesis normal de GDH. Estos datos indican que la GS de E. coli, es capaz de regular la expresión del gene gdhA de K. aerogenes. (Streitcher et al., 1976).

En ese mismo sentido, algunos experimentos indican que los niveles de GS guardan relación con los niveles de enzimas reguladas por represión catabólica y que están involucradas en la degradación de compuestos nitrogenados, tales como L-prolina y L-histidina. Por ejemplo, en condiciones de alta expresión de GS (glucosa + amonio bajo), dichas enzimas son sintetizadas rápidamente en ausencia de AMP cíclico.

Todas las observaciones anteriores indican que la glutamino sintetasa, como entidad proteica, regula la expresión de los genes que codifican tanto para GDH, como para algunas otras enzimas involucradas en el metabolismo nitrogenado de K. aerogenes.

Si bien esto funciona así en K. aerogenes y K. pneumoniae, la situación es diferente en E. coli, en donde los niveles de GDH no se alteran significativamente en condiciones de crecimiento en alta o baja concentración de amonio (15mM o 1 mM respectivamente), no obstante la variación en los niveles de GS (Streitcher et al., 1976).

La falta de regulación de gdhA por parte de GS en E. coli y algunos otros datos relacionados, condujeron a Pahel et al., (1978), a proponer que la región promotor-operador del gene gdhA de E. coli, debería ser diferente a la del de K. aerogenes e insensible a la regulación por nitrógeno.

Sin embargo, otros grupos han reportado que en E. coli, los niveles de GDH varían de acuerdo con los niveles de amonio (Varrichio, F., 1969; Senior, P.J., 1975). Aun más, Senior (1975) ha reportado que los niveles de GDH en E. coli W, son más altos en un cultivo continuo en presencia de concentraciones limitantes de amonio, que en condiciones no limitantes. Esto último va en desacuerdo total con lo reportado por los otros grupos.

Datos no publicados obtenidos en nuestro laboratorio (Becerril, B), indican que células de E. coli, cepas CS520 y RR1, crecidas en condiciones en las que el amonio resulta ser un elemento limitante, presentan una actividad de GDH dos veces mayor que en exceso de amonio. Estos datos están de acuerdo con los publicados por Senior (1975).

Senior (1975) propone en su discusión, que la GDH de E. coli, podría funcionar en la dirección catabólica, en condiciones

de limitación de amonio, proporcionando amonio a partir de glutamato. Si analizamos esta proposición en función de las K_M que presenta esta enzima, comparándolas con las de una GDH esencialmente biosintética, tal como la de Salmonella typhimurium, veremos que la GDH de E. coli podría funcionar in vivo en ambos sentidos.

	K_M (mM)	
	<u>Escherichia coli</u> (Sakamoto <u>et al.</u> , 1975)	<u>Salmonella typhimurium</u> (Coulton y Kapoor, 1973)
α cetoglutarato	0.640	4.000
+ NH 4	1.100	0.290
NADPH	0.040	0.019
glutamato	1.300	30.000
+ NADP	0.042	0.013

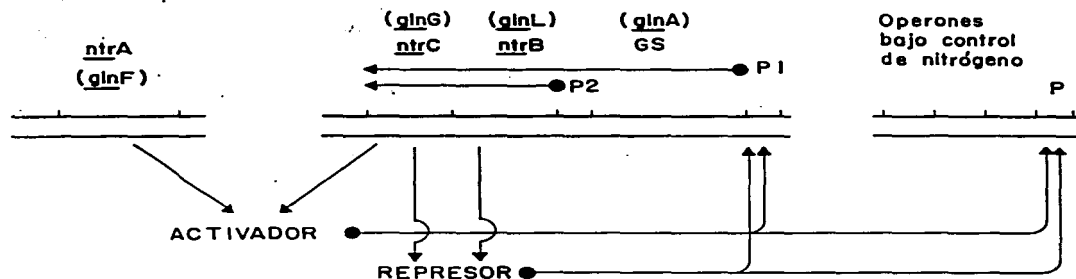
De hecho, existe un reporte en el cual se demuestra que la GDH de E. coli, puede funcionar in vitro en la dirección catabólica con concentraciones fisiológicas de glutamato (Veronesse et al., 1975), a diferencia de la GDH de S. typhimurium, la cual es esencialmente biosintética (Coulton y Kapoor, 1973).

Regresando a la regulación de las vías de asimilación de amonio, es importante mencionar que al igual que los niveles de AMP cíclico afectan la expresión de un buen número de operones involucrados en la asimilación de carbono, recientemente se han publicado una serie de evidencias que apuntan hacia la existencia de un sistema regulatorio que controla la expresión de varios operones involucrados en la asimilación y distribución de nitrógeno en las

enterobacterias. (McNeil et al., 1982; Merrick, 1982). Dicho sistema ha sido llamado Control por Nitrógeno (Merrick, 1982).

Un número considerable de experimentos hechos en S. typhimurium, E. coli, K. pneumoniae y K. aerogenes, han permitido plantear un modelo para explicar el control por nitrógeno (Kustu et al., 1979; Pahel y Tyler, 1979; Magasanik y Stadtman, 1980; McFarland et al., 1981; MacNeil et al., 1982; Merrick, M.J., 1982).

Un esquema que muestra tal modelo es el siguiente:



(donde Y=pirimidina, R= purina y el nucleótido +1, indica el sitio de inicio de la transcripción), puede definir el sitio de reconocimiento para la variante de ARN polimerasa.

El sistema regulatorio ntrB - ntrC, en el cual un sólo operón codifica para un represor y un activador respectivamente, es raro entre los sistemas regulatorios hasta ahora descritos. Solamente se conoce otro ejemplo semejante, y se trata precisamente de otra vía que involucra al nitrógeno. Tal es el caso de la regulación de los genes nif que intervienen en la fijación del nitrógeno atmosférico en K. aerogenes. Dichos genes son regulados por los productos nifL y nifA que actúan como represor y activador respectivamente (Merrick, M.J., 1982).

La semejanza entre los sistemas antes mencionados, va más allá de la simple estructura del operón ya que como se mencionó anteriormente, se ha reportado la existencia de una secuencia conservada en la región regulatoria de los operones que se sabe están controlados por el mecanismo general de activación por ntrA - ntrC. De especial interés es el hecho de que el producto de nifA (activador de los genes nif), puede sustituir al producto del gene ntrC, para activar los promotores de algunos genes involucrados en la asimilación de nitrógeno, incluyendo los promotores de nifL, glnA y del gene de la nitrogenasa (nifH) de Rhizobium meliloti (Ow et al., 1983).

Si bien este panorama de la regulación de las vías de asimilación y distribución de amonio opera sin ambigüedad en S. typhimurium, K. pneumoniae y K. aerogenes, otro parece ser el caso en Escherichia coli, en donde existe controversia al menos de como se lleva a cabo la regulación del gene que codifica para GDH.

Una evaluación crítica de las publicaciones acerca de la regulación del gene gdhA de E. coli, indica que los datos son más bien escasos y contradictorios.

En función de la importancia de la enzima glutamato deshidrogenasa en el metabolismo nitrogenado de Escherichia coli y la atención que recientemente ha adquirido en el campo de la Biotecnología (Windass et al., 1980), una investigación que pretendá conocer el (los) mecanismo(s) de regulación del gene gdhA es necesaria y justificable. Es en este sentido que el presente trabajo se enfoca, ya que forma parte de un proyecto que pretende conocer la regulación del gene estructural que codifica para la enzima GDH de Escherichia coli K-12.

La estrategia inicial que se escogió para desarrollar la investigación fue la siguiente:

- 1.- Aislamiento y clonación del gene estructural y regiones flanqueadoras 5' y 3'.
- 2.- Determinación de la dirección de la transcripción del gene en el fragmento clonado.
- 3.- Secuenciación nucleotídica del gene y de las regiones que lo flanquean.
- 4.- Identificación y caracterización de las regiones de regulación del gene.
- 5.- Estudio de los mecanismos de regulación de la expresión del gene.

Los tres primeros puntos han sido reportados en la literatura por nuestro laboratorio (Covarrubias et al., 1980; Sánchez Pescador et al., 1982; Sanvicente et al., 1983; Valle et al., 1983; Valle et al., 1984). El desarrollo del presente proyecto pretende completar el punto 4 e iniciar los estudios sobre el punto 5.

El estado actual del conocimiento de la zona de regulación 5' del gene gdhA es el siguiente:

El análisis de la secuencia (200 pb) de la región inmediatamente anterior al codón de iniciación (Fig. 1), permitió proponer la existencia de tres posibles promotores por criterios de secuencia consenso (Valle et al., 1983). Estos supuestos promotores se denominaron P1, P2 y P3 en relación al orden que guardan con respecto a la metionina inicial, siendo P1 el más cercano.

Con base en los reportes recientes de secuencias consenso de los promotores de E. coli (Hawley y McClure, 1983; McClure, 1985; Galas et al., 1985), el promotor que más se apega al consenso es el supuesto promotor P3.

Estudios de transcripción in vitro, indican que el fragmento TaqI - HpaII (Fig. 1), une a la ARN polimerasa. El promotor presente en este caso sería P1 (Valle et al., 1983).

Por otro lado, en experimentos in vivo, el fragmento HpaII - HpaII clonado en el sitio de ClaI del pBR327, ensayado en minicélulas, produce un péptido de 63 aminoácidos independientemente de la orientación (Valle et al., 1983). Estos datos indican que este fragmento contiene un promotor propio. Sin embargo, esto no sucede con el fragmento TaqI - HpaII clonado y ensayado de la misma manera, es

Fig. 1. A).- Secuencia nucleotídica de la región correspondiente a los primeros 102 aminoácidos y a la región de control del gene gdhA. El primer nucleótido codificante se designa como +1. También se indica la localización de algunos sitios de restricción. De las posiciones +34 a +54, se presenta la secuencia del oligonucleótido utilizado como "primer" en el mapeo del promotor (ver Materiales y Métodos). B).- Secuencia nucleotídica expandida de la región -192 a +9 en la que se indican las cajas de Pribnow de los supuestos promotores P1, P2 y P3, como líneas gruesas sobre la secuencia. Debajo de ésta, se indican diferentes secuencias repetidas directas como: líneas dobles, punteadas, de círculos llenos, vacíos y compuestas. La flecha de doble punta (-97 a -76), indica la secuencia que conforma un posible sitio de unión a CRP, mientras que las flechas convergentes denotan secuencias palindrómicas. Adaptada de Valle et al., 1983.

-192

-120

HpaII

EcoRII

A

CCGGGTGGCAAAACTTTAGCGTCTCAGGTTATCGCATTGGTTATGAGATTACTCTCGTTATTAATTTGCTTCC

11

TaqI -76 -60 -50

TGGGTCATTTTTTCTTGCTTACCAGTCACATTCTTGATGGTATAGTCGAAAAC TGCAAAAGCACATGACATAAAC

+1 Sau3A

AACATAAGCACAAATCGTATTAATATATAAGGGTTTTATATCT ATG GAT CAG ACA TAT TCT CTG GAG

Ser Phe Leu Asn His Val Gln Lys Arg Asp Pro Asn Gln Thr Glu Phe Ala Gln Ala
TCA TTC CTC AAC CAT GTC CAA AAG CGC GAC CCG AAT CAA ACC GAG TTC GCG CAA GCG

← 3' TTTG GTA CAG GTT TTC GCG CTG 5'

Val Arg Glu Val Met Thr Thr Leu Trp Pro Phe Leu Gln Gln Asn Pro Lys Tyr Arg
GTT CGT GAA GTA ATG ACC ACA CTC TGG CCT TTT CTT GAA CAA AAT CCA AAA TAT CGC

HpaII

Gln Met Ser Leu Leu Glu Arg Leu Val Glu Pro Glu Arg Val Ile Gln Phe Arg Val
CAG ATG TCA TTA CTG GAG CGT CTG GTT GAA CCG GAG CGC GTG ATC CAG TTT CGC GTG

Val Trp Val Asp Asp Arg Asn Gln Ile Gln Val Asn Arg Ala Trp Arg Val Gln Phe
GTA TGC GTT GAT GAT CGC AAC CAG ATA CAG GTC AAC CGT GCA TGG CGT GTG GAG TTC

HpaI

Ser Ser Ala Ile Gly Pro Tyr Lys Gly Gly Met Arg Phe His Pro Ser Val Asn
AGC TCT GCC ATC GGC CCG TAC AAA GGC GGT ATG CGC TTC CAT CCG TCA GTT AAC

B

-192 CCGGGTGGCAAAACTTTAGCGTCTGAGGTTATCGCATTGGTTAT -180

GAGATTACTCTCGTTATTAATTGCTTTCCTGGGTCATTTTTTTC
TaqI

TTGCTTACCAGTCACATTCTTGATGGTATAGTCGAAAAC TGCAAA
P₂

GCACATGACATAAAACAACATAAAGCACAAATCGTATTAATAATATAAG
P₁ P₂ P₁ -80

GGTTTTATATCTATG GAT CAG
+1 Met Asp Gln

decir, no se produce ningún péptido. El fenómeno de que las condiciones de transcripción in vitro puedan ser diferentes in vivo, está reportado en la literatura (Rosenberg y Court, 1979).

Estudios preliminares de mapeo por digestión con nucleasa S1, indican que bajo condiciones de crecimiento en alta concentración de amonio, la transcripción se inicia en el supuesto promotor 3 (Valle et al., 1983).

Los datos anteriores sugieren que la región comprendida entre HpaII - TaqI, es esencial para la expresión del gene.

Además de los tres posibles promotores, dentro de la región regulatoria, se detectan una serie de secuencias repetidas directas que podrían ser sitios de interacción con proteínas regulatorias y/o dominios estructurales que le confirieran una arquitectura especial al ADN y que ésta tuviera alguna influencia sobre la regulación del gene.

En esta misma región, entre los promotores 3 y 2, se encuentra un posible sitio de interacción con CRP (CAP, catabolite gene activating protein). La participación de CRP en la regulación de gdhA está documentada en la literatura (Prussiner et al., 1972).

En condiciones de crecimiento en alta concentración de amonio, todo parece indicar que el promotor 3 es funcional. La ubicación del sitio de CRP sobrelapado con el -35 de este supuesto promotor, indica que por lo menos en esa condición metabólica, la activación por CRP no es clara ya que no se han detectado transcritos que inicien en P2 o P1. Sin embargo, en otras condiciones metabólicas P2 o P1 podrían ser activados por dicha proteína.

Otra característica interesante de este gene es, que carece de las señales típicas de atenuación de la transcripción presentes en varios operones de biosíntesis de aminoácidos (Yanofsky, C., 1981).

OBJETIVOS

En el presente trabajo, se planea caracterizar las regiones regulatorias 5' y 3' del gene gdhA. En lo que respecta a la región regulatoria 5' se pretende mapear el (los) promotor(es) que funciona(n) bajo determinadas condiciones metabólicas. De la región regulatoria 3' se planea conocer de que manera participan las secuencias nucleotídicas con simetría de d'ada presentes en esa región flanqueadora en la regulación del gene.

MATERIALES Y METODOS.

El material utilizado en este trabajo, se describe en las publicaciones que se anexan. Los métodos que se enlistan también han sido descritos previamente en esas publicaciones.

- Digestión de ácidos nucleicos con enzimas de restricción.
- Clonación de fragmentos de restricción.
- Crecimiento de diferentes cepas en diferentes medios.
- Secuenciación de ADN en fago M13. Método de Sanger.
- Medición de la actividad específica de GDH.
- Determinación del número relativo de copias de plásmido.

Aquellos métodos no incluidos en las publicaciones mencionadas, se detallan a continuación:

a).- Purificación de ARN total. El método seguido en esta parte fue proporcionado amablemente por la Dra. Alejandra Covarrubias del Centro de Investigación sobre Fijación de Nitrógeno, U.N.A.M., Cuernavaca, Morelos.

Mataraces con 50 ml. de medio "NN" suplementado con 1 mM o 15 mM de NH_4Cl como fuente de nitrógeno y glucosa 11 mM o glicerol 22 mM como fuente de carbono, fueron inoculados con una alícuota de un cultivo de la noche anterior. La densidad óptica inicial se ajustó a 0.1 leída a 620 nm.

Las células fueron colectadas en la fase exponencial temprana. Se cosecharon alícuotas de 25 ml en un tubo Corex, centrifugando 30 segundos a 10,000 rpm en un rotor JA20 (Beckman Instruments), deteniendo el rotor lo más rápidamente posible.

Se decantó el sobrenadante y se colocó el tubo inmediatamente en un baño de hielo seco.

El botón de células se resuspendió en 3 ml de solución de lisis (acetato de sodio 20 mM pH 5.5, EDTA 1 mM, SDS 0.5%), valiéndose de una pipeta Pasteur invertida. Se prosiguió la lisis calentando a 70 °C durante 40-60 segundos en baño María.

Transcurrido el tiempo de la lisis, se añadió inmediatamente un volumen (3 ml) de fenol saturado (NaOAc 20 mM pH 5.5, EDTA 1mM), previamente calentado a 65°C .

Se mezclaron las fases en un baño con agitación (65°C, 300 rpm), durante 10 minutos.

Se centrifugó a 8,000 rpm durante 5' para separar las fases, en un rotor JS13 (Beckman Instruments).

Se pasó el sobrenadante a otro tubo Corex y se repitió la extracción con fenol, dos veces más.

La fase acuosa de la última extracción se precipitó con 0.1 volúmenes de NaOAc 3M pH 5.5 y 2.5 volúmenes de etanol absoluto, dejando precipitar el ARN de 2 horas a toda la noche a -20°C.

Se centrifugó a 8,000 rpm durante 45 minutos en un rotor JS13 y se decantó el sobrenadante.

El ARN se secó en una cámara de vacío o con una corriente suave de aire filtrado.

El precipitado de ARN, se resuspendió en 500 µl de agua tratada con dietil pirocarbonato (DPC) al 0.07% V/V. La concentración, y la aparente pureza de la preparación, se determinaron mediante el uso de una dilución adecuada y la obtención de la relación de absorbancias a 260 y 280 nm. Una unidad óptica de absorbancia a 260 nm equivale a una concentración de 40 µg/ml.

La integridad del ARN de las preparaciones anteriores, se determinó visualmente, corriendo una muestra del ARN en un gel de agarosa al 1.5% con amortiguador E (borato de sodio 5mM, sulfato de sodio 10 mM, ácido bórico 5 mM pH 8.2), conteniendo 5 mM de hidróxido de metil-mercurio (CH₃HgOH).

5-10 μg de ARN, se llevaron a un volumen de 15 μl con amortiguador E y se adicionó un volumen igual de CH_3HgOH 10 mM, glicerol 10%, azul de bromofenol 0.5%.

El gel se corrió a 30 Volts (50 mA), durante 14 horas a temperatura ambiente con recirculación del amortiguador, en una campana de extracción.

El gel se tiñó sumergiéndolo 30' en una solución que contenía NH_4Cl 0.5N y después 10' en 2 mercaptoetanol 5 mM, bromuro de etidio 0.5 $\mu\text{g}/\text{ml}$.

Se iluminó el gel en un transiluminador de luz UV y se hizo una inspección visual de la integridad de los ARN ribosomales 23, 16 y 5S.

b).- Marcaje de oligonucleótidos con ^{32}P . Este método también fue proporcionado amablemente por la Dra. Alejandra Covarrubias.

Para marcar el oligonucleótido se usó la siguiente mezcla de reacción:

15 picomoles de oligonucleótido (3.5 pmoles/ μl = 4.3 μl)
 5 μl de amortiguador para Cinasa del fago T4 (Tris-HCl 50 mM, pH 7.6, MgCl_2 10 mM, 2 mercaptoetanol 10 mM)
 10 μl de [$\gamma^{32}\text{P}$] ATP (Amersham, 7000 Ci/mmol, 10 $\mu\text{Ci}/\mu\text{l}$ = 100 μCi)
 5 μl de Cinasa del fago T4 (Amersham, 3 U/ μl)
 25.7 μl de agua, para un total de 50 μl .

Se incubó la reacción una hora a 37°C y se inactivó la enzima calentando la mezcla de reacción a 65°C durante 10 minutos.

La eficiencia relativa del marcaje y la integridad del oligonucleótido se determinaron cuando se corrió una alícuota de la mezcla de reacción en un gel de acrilamida al 20% y urea 8 M, usando como marcador un oligonucleótido de tamaño conocido igualmente marcado. Se cubrió el gel con una película sensible a los rayos X, se puso a exponer en una carpeta radiográfica con pantalla intensificadora, durante 2 horas, al cabo de las cuales la película fue revelada.

c).- Extensión de primeró. El protocolo seguido fue el reportado por León et al (1985), con pequeñas modificaciones.

Para la implementación de esta metodología, fue necesaria la síntesis química de un oligonucleótido de ADN, el cual se diseñó de tal manera que fuera complementario a una secuencia del ARN mensajero específico de GDH, distante en aproximadamente 90 nucleótidos del supuesto extremo 5' terminal (ver Fig. 1). Este oligonucleótido fue sintetizado por la Biol. Laura Riba, siguiendo el método del triéster en fase sólida, bajo la asesoría del QFB Mario Alberto Cuevas en la Unidad de Síntesis Química de Macromoléculas del Centro de Investigación sobre Ingeniería Genética y Biotecnología, U.N.A.M.

El nucleótido así diseñado fue un 21ámero con la secuencia:

5' G T C G C G C T T T T G G A C A T G G T T 3'

El protocolo seguido para la técnica de extensión de primeró fue el siguiente:

50-100 µg de ARN total, se mezclaron con 0.1 - 1 pmole de oligonucleótido marcado con ³²P ATP, en un tubo Eppendorf de 1.5 ml.

La mezcla se precipitó con 0.1 volúmenes de NaOAc 3M y 2.5 vols. de etanol absoluto. Se puso 10-20' en hielo seco.

Se centrifugó a 12,000 rpm durante 15' en una microcentrífuga Beckman.

Se descartó el sobrenadante y se secó el precipitado en una cámara con vacío.

Se resuspendió en 1 μ l de Tris-HCl 50 mM, pH 8.3, 2 μ l EDTA 1 mM y 2.7 μ l de agua.

Se hirvió durante 5' y se congeló inmediatamente en hielo seco.

Se descongeló a temperatura ambiente durante 45 minutos.

Se incubó en baño María a 43°C durante 3 horas y se agregaron:

1 μ l Tris-HCl 1M, pH 8.3

2 μ l MgCl₂ 0.1M

1 μ l dTTP 20 mM

1 μ l dCTP 20 mM

1 μ l dGTP 20 mM

1 μ l dATP 20 mM

0.8 μ l KCl 1M

0.4 μ l DTT (ditiotretol), 0.2M

1-3 μ l agua

15-25 U transcriptasa reversa (3-5 μ l, 5 U/ μ l, Prom:

Biotec).

Se incubó en baño María a 43°C durante 30-45 minutos.

Se hirvió durante 5 minutos.

Se agregó 1 μ l de ARNasa A (Sigma Chemical, Co., 10 mg/ml), se llevó a 100 μ l con agua y se incubó 1 hora a 37°C.

Se extrajo con un volumen de fenol saturado con agua.

Se recuperó la fase acuosa y se agregaron 5 μ l de ARNt (1 μ g/ μ l).

Se precipitó con 0.1 volúmenes de NaOAc 3M y 2.5 vols. de etanol.

Se lavó 2 veces con 250 μ l de etanol 70%.

Se secó, se resuspendió en 3 μ l de agua y se agregaron 5 μ l de mezcla de detención (stop mix), de secuencia (formamida 90%, xilencianol 0.02%, azul de bromofenol 0.02%)

Se hirvió 2-3 minutos, se puso inmediatamente en un baño de hielo de agua y se corrieron las muestras en un gel de secuencia (acrilamida al 6%, urea 8M).

En ambos extremos del gel se corrieron secuencias nucleotídicas conocidas, para determinar con precisión el tamaño de las bandas del ADNc.

d).- Mutagénesis de la región regulatoria. El desarrollo de esta metodología, es una contribución original del autor de esta tesis.

Las manipulaciones de la región regulatoria se hicieron en el plásmido pSAE422 (Sánchez-Pescador et al., 1982), el cual es un

derivado del pBR322. A este último, se le cambió la región comprendida entre PstI - EcoRI, por un fragmento de tamaño equivalente, que porta la región regulatoria del gene gdhA y la secuencia que codifica para los primeros 102 aminoácidos de la GDH (Fig. 2).

Las razones para usar el plásmido pSAE422, fueron por un lado que su secuencia nucleotídica es conocida y por otro, que los patrones de restricción fueron los adecuados para distinguir rápidamente las clonas deseadas.

Las manipulaciones fueron las siguientes:

1.- Se tomó ventaja de la presencia de una secuencia reconocida por la endonucleasa de restricción TaqI (T'CGA, Fig.1), en la cual el sitio de corte involucra a la T de la posición -8 de la región de -10 (caja de Pribnow), del supuesto promotor 3.

Se hizo una digestión parcial del plásmido con TaqI, se purificó la banda de plásmido lineal y se puso a polimerizar en presencia de los cuatro deoxinucleótidos y el fragmento Klenow de la ADN polimerasa.

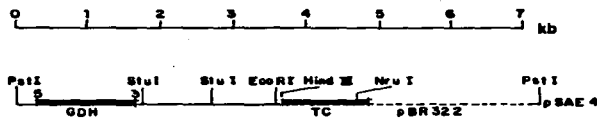
Después de polimerizar el plásmido, se ligó a sí mismo y se usó para transformar la cepa PA340 (gdh-1, gltB31).

El resultado de esta manipulación fue la introducción de un par de bases entre el -10 del supuesto promotor y el sitio de inicio de la transcripción, al cambiar el sitio de TaqI (T'CGA), por un sitio de NruI (TCG'CGA). El plásmido generado se llamó pSAE430 (Fig.2).

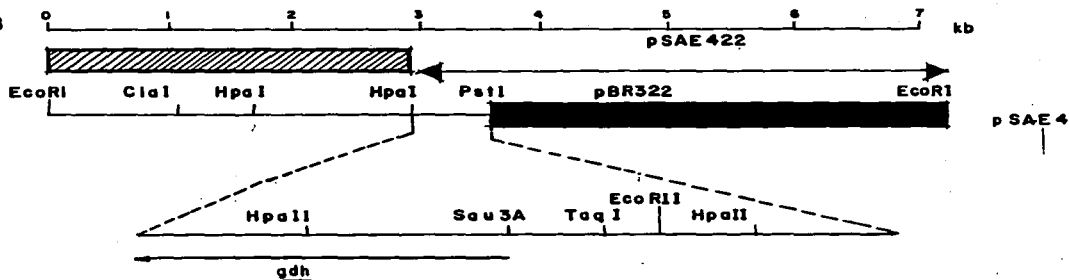
La selección de la clona esperada se basó en la búsqueda, entre las resistentes a tetraciclina, de aquella que tuviera dos bandas de

Fig. 2. Mutagénesis por inserción de oligonucleótidos de la región regulatoria del gene gdhA. Los detalles experimentales están dados en Materiales y Métodos. La figura muestra un mapa físico del plásmido pSAE4 linearizado con PstI (A), en donde se indican algunos sitios de restricción. Las líneas gruesas indican los genes de gdhA y Tc^R. En este último se indica el sitio de NruI. Un mapa físico del plásmido linearizado con EcoRI se muestra en (B), en donde se esquematiza la construcción del plásmido pSAE422 (Sánchez-Pescador et al., 1982). El área sombreada representa la región deletada durante dicha construcción. La flecha con doble punta representa la región correspondiente al plásmido pSAE422 y la línea gruesa al vehículo pBR322. Debajo del mapa, una amplificación de la región regulatoria y la región que codifica para los primeros 102 aminoácidos. La flecha indica la dirección y el sitio de inicio de la traducción. Un esquema de las manipulaciones de la región regulatoria se muestran en (C), en donde se representa únicamente la región HpaII-HpaII y los sitios manipulados (TaqI y EcoRII). Los números entre los sitios de restricción indican la distancia entre ellos en pares de bases. La flecha indica lo mismo que en (B). Debajo de los mapas de restricción se indican las secuencias nucleotídicas relevantes. Se encuentran subrayados los sitios de TaqI (pSAE422), NruI (pSAE430), mono "linker" de EcoRI (pSAE432) y el di "linker" (pSAE433). Las flechas verticales indican el sitio de inicio de la transcripción determinada en los experimentos de extensión de primer (ver Materiales y Métodos, y Resultados).

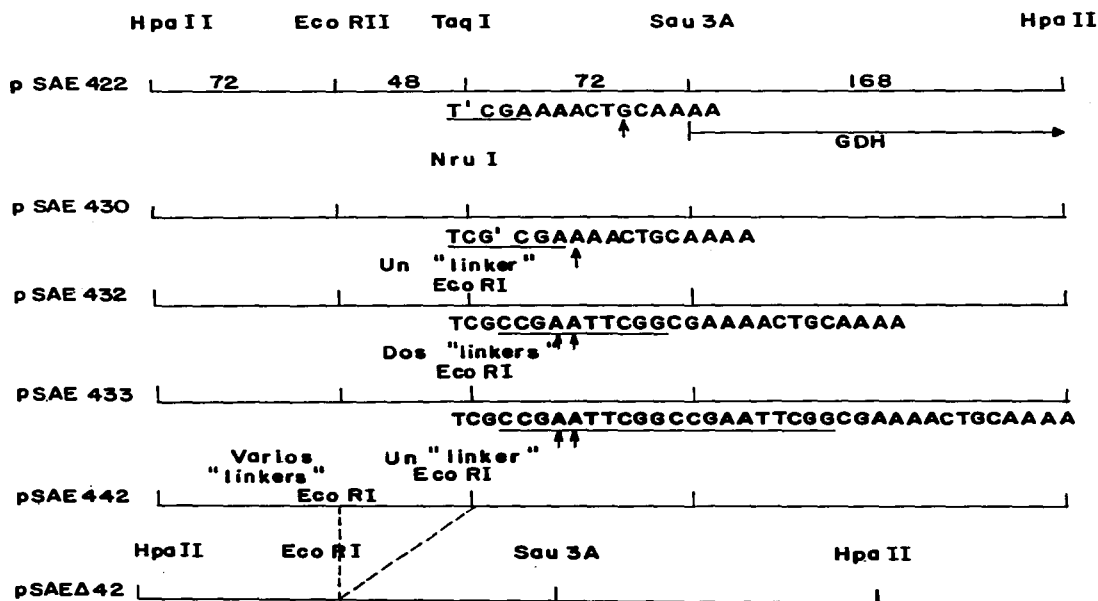
A



B



C



NruI, una de 2850 y otra de 1350 pb que son los tamaños esperados en función de la presencia de un sitio único de esta enzima a 1350 pb del sitio original de TaqI de la región regulatoria del gene gdhA en el plásmido pSAE422. Dicho sitio único se localiza en la coordenada 972 del plásmido pBR322, quedando dentro de la región que codifica para la resistencia a tetraciclina (Fig.2A).

2.- La siguiente manipulación consistió en la modificación del nuevo sitio de NruI, mediante la incorporación en él, de un adaptador sintético de ADN, portador de un sitio de reconocimiento para la endonucleasa de restricción EcoRI. Este adaptador fue amablemente proporcionado por el Dr. Xavier Soberón (Departamento de Genética y Biología Molecular, Centro de Investigación sobre Ingeniería Genética y Biotecnología, U.N.A.M., Cuernavaca, Morelos).

Para realizar dicha manipulación, se hizo una digestión parcial del plásmido pSAE430 con NruI. Se purificó la banda lineal y se puso a ligar en presencia de un exceso de "linkers" de EcoRI. El producto de la reacción anterior se usó para transformar la cepa HB101.

La selección de la clona deseada, se basó en la búsqueda entre las resistentes a tetraciclina, de aquella que tuviera un sólo sitio de NruI y que tuviera un sitio de EcoRI, a 1350 pb del de NruI.

Para estar seguros de que estuviera presente un solo "linker" de EcoRI, se purificó el ADN del plásmido esperado y se digirió extensivamente con EcoRI, se purificó la banda lineal, se ligó, y se usó para transformar la cepa PA340.

Las clonas del plásmido esperado, que portan más de una molécula del adaptador sintético ("linker"), se pueden distinguir porque 2

moléculas del adaptador que se ligan contiguamente, generan un sitio adicional de corte para la endonucleasa de restricción HaeIII. Basándose en este argumento, se seleccionó una de estas clonas, se purificó el ADN del plásmido y se usó para transformar la cepa PA340.

Al plásmido portador de una sola molécula de adaptador se le llamó pSAE432 y al que porta más de una, se le llamó pSAE433 (Fig.2).

3.- El objetivo de la siguiente manipulación consistió en insertar una molécula del adaptador en el sitio de Eco RII, distante 50 pb hacia arriba del sitio nuevo de EcoRI en el plásmido pSAE432.

La estrategia consistió en hacer una digestión parcial del plásmido pSAE432 con la endonucleasa de restricción BstNI, la cual es un isoesquizómero de EcoRII.

La razón de usar BstNI en lugar de EcoRII, fue que esta última requiere que el sitio que reconoce esté metilado, haciendo necesaria la purificación del plásmido de una cepa que metile los sitios de EcoRII en el plásmido.

Una vez obtenida la digestión parcial, se purificó la banda lineal y se puso a polimerizar en presencia del fragmento Klenow de la ADN polimerasa y los cuatro deoxinucleótidos.

El producto de la polimerización se puso a ligar en presencia de un exceso de "linkers" de EcoRI. Esta mezcla de ligamiento, se usó para transformar la cepa PA340.

La selección de la clona esperada consistió en localizar de entre las resistentes a tetraciclina, a aquella que diera una banda de 50 pb, por efecto de la digestión con EcoRI. El plásmido seleccionado fue

llamado pSAE442.

4.- La última manipulación consistió en eliminar la región comprendida entre EcoRII y TaqI de la región regulatoria del gene gdhA, es decir, la secuencia que contiene al supuesto promotor P3.

La estrategia para realizar esta manipulación consistió simplemente, en digerir extensivamente al plásmido pSAE442 con EcoRI, purificar la banda grande y ponerla a ligar contra sí misma. El producto de la mezcla de ligamiento se usó para transformar la cepa PA340.

La selección de la clona se limitó a localizar entre las resistentes a tetraciclina, a aquella que tuviera un sólo sitio de EcoRI. El plásmido así seleccionado fue llamado pSAE Δ 42.

RESULTADOS.

Los resultados serán presentados en dos partes, una en lo referente a la región flanqueadora 5' del gene y la otra correspondiente a la región flanqueadora 3', haciendo entre ambas, una breve referencia a la parte estructural del gene.

Las estrategias iniciales para secuenciar la región regulatoria 5' del gene gdhA, y la región que codifica para los primeros 102 aminoácidos de la enzima GDH de Escherichia coli, han sido reportados con anterioridad (Valle et al., 1983); también se han reportado la subclonación de diversos fragmentos portadores de distintas partes de la región regulatoria 5' y el análisis de la capacidad de dichos fragmentos para promover la expresión de éste y otros genes in vitro e in vivo (Valle et al., 1983; Valle, 1984).

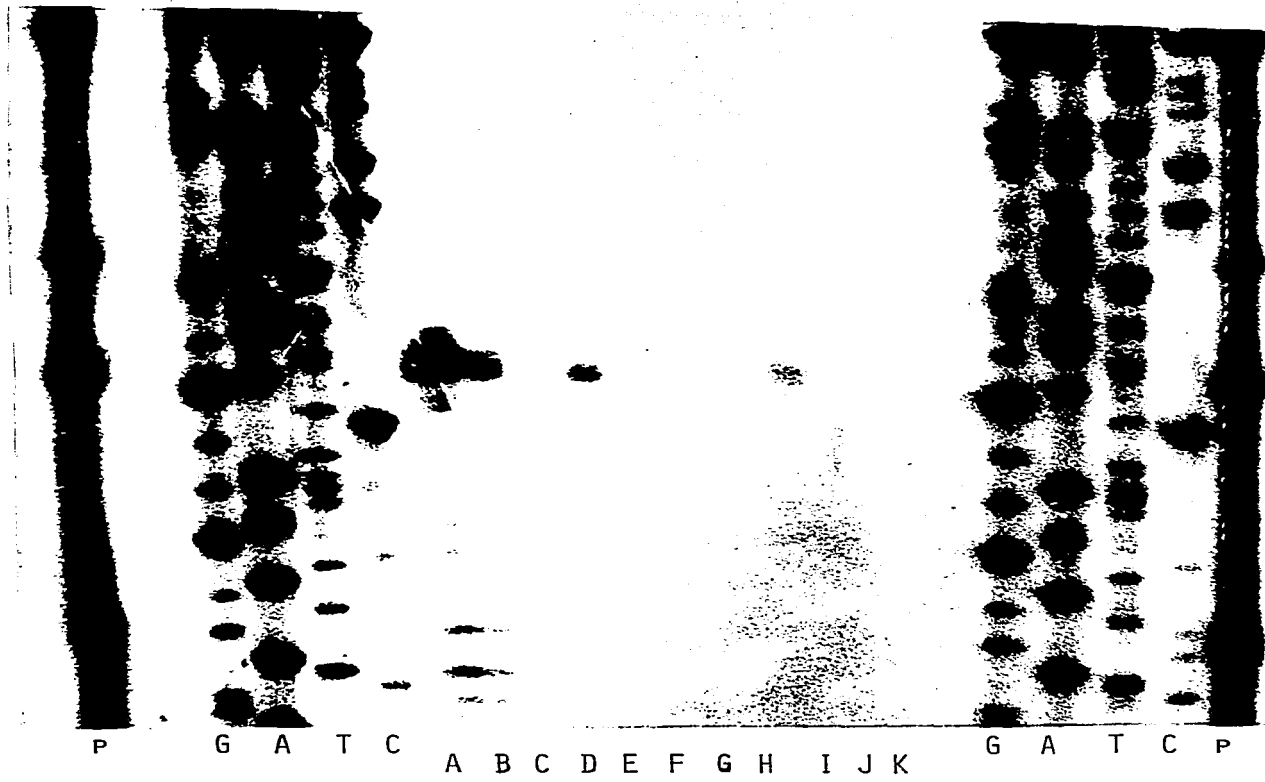
Región regulatoria 5' (Mapeo del promotor).

En base al análisis de la secuencia nucleotídica de la región regulatoria y los experimentos iniciales encaminados a mapear el promotor, la descripción de los cuales ha sido mencionada en la introducción, decidí realizar tal mapeo, de manera fina. Para cumplir con este objetivo, se hizo uso de la metodología de extensión de "primer" (ver Materiales y Métodos).

Las condiciones iniciales bajo las cuales se mapeó el promotor fueron a partir de ARN total extraído de las cepas CS520 (*gdhA+*) y PA340 (*gdhA-*), transformada con el plásmido pSAE4(*gdhA+*, Sánchez-Pescador *et al.*, 1982)(Fig. 2), crecidas en medio mínimo "NN" suplementado con glucosa (11 mM) o glicerol (22 mM) como fuente de carbono y NH_4Cl 1 mM o 15 mM como fuente de nitrógeno.

Los resultados de este experimento se muestran en la figura 3. Los primeros cuatro y los últimos cuatro carriles, marcados como "GATC", muestran una secuencia nucleotídica conocida, la cual sirve para precisar el tamaño de las bandas de los ADNc, en los carriles experimentales (A-H). I-K, son carriles de control. El punto preciso de inicio fue localizado a partir de la determinación del tamaño de la banda mayoritaria. Como se conoce el punto en el que se inicia la extensión del oligonucleótido que sirve como primero (en este caso el 21ámero descrito en Materiales y Métodos), y se puede precisar el tamaño de la banda mayoritaria antes mencionada, automáticamente se conoce el sitio en donde la transcriptasa reversa deja de incorporar nucleótidos; este punto corresponde al extremo 3' del mensajero de la GDH. En la vecindad de 5-8 nucleótidos hacia arriba de este punto en la secuencia, debe encontrarse el -10 del promotor (Hawley y McClure,

Fig. 3. Mapeo del promotor del gene gdhA. Los detalles experimentales están dados en Materiales y Métodos. Esta figura muestra los resultados de la extensión de "primer" usando ARN extraído de las cepas PA340/pSAE4 y CS520 crecidas en diferentes condiciones metabólicas. El carril denominado p, corresponde al plásmido pBR322 digerido con HpaII y marcado con ³²P. Se muestran las bandas de 110 y 122 pb. Los carriles marcados como GATC representan la secuencia nucleotídica del fago M13mp18 utilizada como marcador de peso molecular. El carril A corresponde al ADNc sintetizado al usar como molde ARN extraído de la cepa PA340/pSAE4 crecida en medio "NN" suplementado con glucosa y NH₄Cl 15 mM. Carril B, glucosa y NH₄Cl 1 mM. Carril C, glicerol y NH₄Cl 15 mM. Carril D, glicerol y NH₄Cl 1 mM. Los carriles E-H, muestran los ADNc respectivos de la cepa CS520. Después de una corrección de 2 bases en todos los casos, la banda más intensa ubica el inicio de la transcripción en la posición -63 de la figura 1. Los carriles I-K, representan controles internos. Carril I, condiciones del carril A, excepto que el ARN total es reemplazado por ARN de transferencia. Carril J, no contiene "primer". Carril K, no lleva ARN de ningún tipo.



P

G

A

T

C

A

B

C

D

E

F

G

H

I

J

K

G

A

T

C

P

122

110

1983).

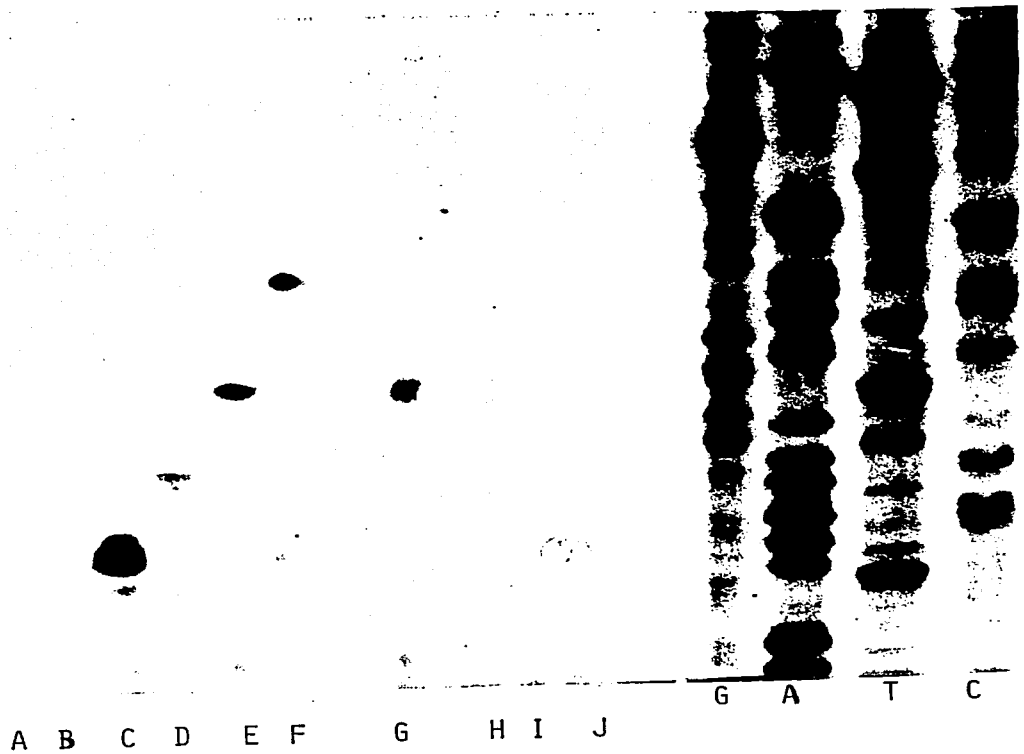
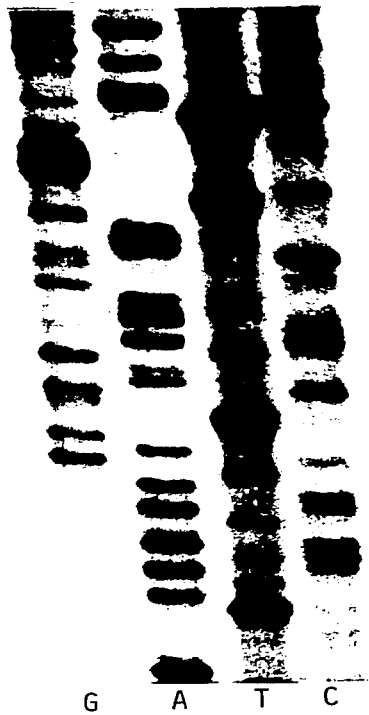
Como se puede ver en esta figura, en todas las condiciones metabólicas probadas se produce una banda mayoritaria, la cual corresponde, en base a las secuencias puestas de ambos lados, a un tamaño de 117 nucleótidos. Tomando en consideración que el oligonucleótido que se usó como primero es un 21ámero, los 96 nucleótidos que fueron incorporados en el ADNc, permiten ubicar el inicio de la transcripción en el nucleótido -63 (B), dado que el extremo 3'OH del oligonucleótido se ubica en la posición +32. La numeración toma como referencia la A del triplete ATG, la cual es considerada +1. Hacia arriba de esta A la numeración es - n y hacia abajo + n, donde n es cualquier número (Fig.1).

Los datos presentados en la figura 3, permiten ubicar el inicio de la transcripción enfrente del promotor 3. Es interesante hacer notar que tanto en condiciones de crecimiento en glucosa o glicerol como fuente de carbono y en exceso (15 mM) o limitante (1 mM) el amonio, se use el mismo promotor tanto en la cepa silvestre (CS520), como en la cepa con plásmido (PA340/pSAE4), para promover el gene gdhA.

Para estar seguros de que el promotor 3 es el único funcional al menos en la condición de glucosa-alto amonio, se realizaron una serie de manipulaciones en dicho promotor (ver Materiales y Métodos). Se purificó el ARN total de las cepas portadoras de los plásmidos producto de las manipulaciones y se hizo extensión de primero.

Los resultados de esas manipulaciones se presentan en la figura 4. Como se puede observar, al igual que en la figura 3, en ambos

Fig. 4. Extensión de "primer" usando ARN total de las cepas que portan plásmidos con modificaciones de la región regulatoria del gene gdhA. Las indicaciones relacionadas con esta figura, son las mismas que para la Fig. 3. Carril A, corresponde al ADNc sintetizado a partir del ARN extraído de la cepa que porta el plásmido pSAE442. Carril B, pSAE442. Carril C, pSAE422. Carril D, pSAE430. Carril E, pSAE432. Carril F, pSAE433. Carril G, pSAE432; es este caso la muestra fue tratada con NaOH en lugar de ARNasa (ver Materiales y Métodos). Los carriles H-J son carriles de control como en la figura 3. Los detalles experimentales están dados en Materiales y Métodos y en la figura 2.



extremos de los carriles experimentales (A-G), y los carriles de control (H-J), se encuentran los carriles que permiten leer una secuencia nucleotídica conocida y así poder precisar el tamaño de las bandas mayoritarias.

En el carril C, se observa una banda mayoritaria de 117 nucleótidos que corresponde al ADNc obtenido al usar como molde el transcrito original, proveniente del plásmido pSAE422. La transcripción en este caso se inicia en el nucleótido -63 (G) (Fig. 1). El carril D, muestra que el ADNc migra como si tuviera 5-6 nucleótidos más. En este caso el transcrito proviene del plásmido pSAE430, el cual tiene una inserción de un par de bases entre la posición de -7 del promotor y el inicio original de la transcripción (Fig. 2). El tamaño de este nuevo ADNc, ubica el nuevo sitio de inicio de la transcripción en el trecho de As que se encuentran inmediatamente hacia arriba del inicio original. El nuevo inicio podría ser la primera o la segunda de esas Adeninas (ver Fig. 2).

En el carril E, se muestra el tamaño del ADNc producido al usar como templado el transcrito sintetizado a partir del plásmido pSAE432. En este caso, la diferencia con respecto al plásmido anterior (pSAE430), es la inserción de una molécula de ADN sintético de 10 pb. La diferencia de tamaño obtenida, fue exactamente de 10 nucleótidos con respecto al ADNc del carril anterior. En función de la secuencia del decanucleótido usado en esta inserción (5' C C G A A T T C G G 3'), el nuevo sitio de inicio de la transcripción es ahora cualquiera de las dos As del decanucleótido.

El siguiente carril (F), corresponde al ADNc sintetizado a partir del transcrito proveniente del plásmido pSAE433, el cual porta una

molécula adicional del deca-nucleótido arriba mencionado, es decir en lugar de una, tiene 2 moléculas adyacentes.

La predicción en este caso es que si la transcripción se inicia en cualquiera de las As del decanucleótido, entonces la diferencia de tamaño con respecto al carril anterior debe ser de 10 nucleótidos que es el tamaño de la secuencia interpuesta con relación al nuevo inicio, es decir, una molécula del decanucleótido. La predicción es cierta dado que la diferencia de tamaños es efectivamente 10 nucleótidos entre el ADNc que se muestra en este carril con respecto al anterior.

En el carril B, se muestra el resultado de la extensión de primero realizada tomando como molde el transcrito producido por el plásmido pSAE Δ 42. En este caso no se observa ninguna banda, lo cual era de esperarse dado que el plásmido tiene eliminado el promotor (ver Materiales y Métodos), y por tanto no se produce transcrito específico.

En el carril A, se muestra el resultado obtenido con el transcrito producido por el plásmido pSAE442, el cual además de portar un "linker" de EcoRI en la región de -10 del promotor, también porta varios "linkers" iguales en el sitio de EcoRII de la región regulatoria (Figuras 1 y 2). La predicción en esta ocasión, es que si la región aledaña al sitio de EcoRII, no tiene influencia sobre la expresión del gene, el ADNc producido debe ser del mismo tamaño que el del plásmido pSAE432, o sea el que se usó para generar el pSAE442 (ver Materiales y Métodos). No obstante que se produce el ADNc del tamaño predicho, la intensidad de la banda es muy inferior a la esperada, lo cual podría indicar que la manipulación del sitio de EcoRII, pueda tener algún efecto sobre la expresión del gene gdhA.

En resumen los resultados obtenidos con la implementación de la técnica de extensión de primer, permiten asegurar que al menos en las condiciones metabólicas probadas la transcripción del gene gdhA, es iniciada enfrente del promotor 3. Esta afirmación se deriva del hecho de que independiente de la manipulación que se haga hacia abajo de la caja de Pribnow (region de - 10), la transcripción siempre se inicia dentro de los 5-8 nucleótidos hacia abajo de la T de dicha región de control. Esta observación está de acuerdo con lo que se ha encontrado tanto in vivo como in vitro para un número importante de genes (Hawley y McClure, 1983), es decir, el sitio de inicio de la transcripción se encuentra dentro de los 5-8 nucleótidos hacia abajo de T de la posición -8 de la caja de Pribnow.

Region estructural del gene gdhA.

a).- Secuencia nucleotídica y de aminoácidos

Un mapa de restricción fino del gene gdhA de Escherichia coli K-12, se presenta en la figura 1 de la publicación correspondiente (Valle et al., 1984). El fragmento ClaI-PstI del plásmido pSAE4, fue digerido con las enzimas de restricción HaeIII, HpaII, o Sau3A, y los fragmentos de ADN obtenidos se clonaron en el fago M13mp7 en el sitio de ClaI (fragmentos generados por HpaII y TaqI), en BamHI (fragmentos generados por Sau3A), o en SmaI (fragmentos generados por HaeIII). Los fagos recombinantes fueron utilizados para determinar la secuencia del ADN, por el método de Sanger, según los procedimientos reportados por Heidecker et al. (1981) y Messing et al. (1980).

En la figura 2 de esta misma publicación (Valle et al., 1984), se presenta la secuencia nucleotídica del gene estructural y sus regiones

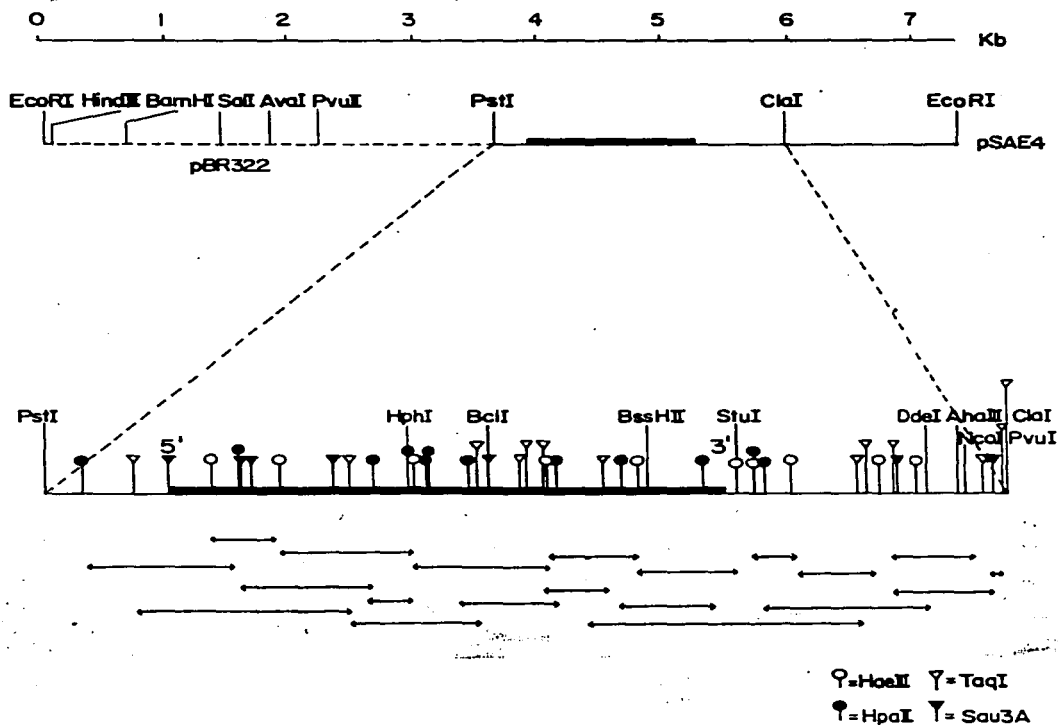


Fig. 1. Physical map of the linear form (*EcoRI* cut) of plasmid pSAE4. The horizontal dashed line represents the vector pBR322, and the thick bar represents the location of the *gdhA* gene with 3' and 5' ends indicated at the lower part. Different DNA fragments from the plasmid were subjected to sequence analysis with the strategy indicated in the lower portion of the figure.

```

-390
CGEGGTGGCA AACTTTAEC ECTGAGETT ATCGCAATTI GGGTATGAGA TTACTCTGCT TATTAAATTT
-400
CTTTCGECG TCATTTTTTI CTGCTCTACC GTCACTTCTT TCAATGATATA GTCAAGAACT GCAAAAGCAC
-50
ATGCACTAAA CACATARAC ACAATGATAT TAATATATA GGGTTTTATA TCT TAT GAT GAG ACA
          Met Asp Glu Met
100
TAT TGT CTG GAG TCA TGC CTC AAC GAT GTG GAA AAG CCG GAG CCG GAT GAA ACC GAG TTT
Tyr Ser Leu Glu Ser Phe Leu Asn His Val Thr Lys Arg Asp Phe Asn His Thr Thr Glu Phe
150
GCG GAA CCG GTT CGT GAA GTA ATG ACC GAA CTC GCG CCG TTT CTT GAA GAA GAT CCA AAA
Glu Ala Glu Val Arg Glu Val Met Thr Thr Leu Trp Phe Phe Leu Glu Glu Asn Phe Lys
200
TAT CCG CAG ATG TCA TTA CTC GAG CGT CTG GTT GAA CCG GAG CCG GTC ATC GAG TTT CCG
Tyr Arg Glu Met Ser Leu Leu Glu Arg Leu Val Glu Phe Glu Arg Thr Thr Glu Phe Arg
250
GTC GTA TGG GTT GAT GAT CCG AAC CAG ATA CAG GTC AAC CTT GCA TGG CST GTC CAG TTC
Val Val Trp Val Asp Asp Arg Asn Glu Thr Glu Val Thr Arg Ala Trp Arg Val Thr Phe
300
AAC TCT CCG ATC CCG CCG TAA AAA CCG GGT ATG CCG TTC CAT CCG TCA GTT GAG CTT TCC
Ser Ser Ala Thr Cys Phe Thr Lys Glu Cys Met Arg Phe His Phe Ser Val Asn Leu Ser
350
ATT CTC AAA TTC CTC CCG TTT GAA GAA ACC TTC AAA AAT CCG CTG ACT GTC CCG CCG AIG
Ile Leu Lys Phe Leu Thr Phe Leu Thr Phe Phe Lys Asn Ala Leu Thr Thr Leu Phe Met
400
AAC CCG GAT AAA CCG CCG AAC GAT TAC GAT CCG AAA GAA GAA AAC GAA Ser GAA GTC AGC
Asp Cys Glu Lys Glu Thr Ser Asp Phe Asp Phe Lys Glu Lys Ser Thr Glu Val Thr Met
450
CGT TTT TGC CAG CCG CTG ATG ACT GAA CTG TAT CCG CAC CTG CCG GAT ACC GAG CTT
Arg Phe Cys Glu Ala Leu Met Thr Glu Leu Trp Thr Arg His Leu Thr Asp Thr Asp Val
500
CCG GAA CCG GAT GAT CCG GTT CCG GGT GGT GAA GTC CCG TTT ATG CCG CCG ATG ATA AAA
Glu Ala Glu Ser Asp Ile Ser Glu Thr Glu Thr Met Ile Thr Met Ile Thr Met Lys
550
AAC CTC TCC AAC AAT CCG CCG TGC CTC TTC ACC CCG AAG CCG CTT TCA TTT CCG CCG APT
Glu Cys CCG Asn Thr Glu Thr Glu Thr Phe Thr Lys Glu Leu Thr Thr Thr Thr Thr Thr
600
CTT ATT CCG CCG GAA CCG ACC CCG TAC CCG GTT CTC GTT TAT TTC ACA GAA GAA ATA GTA AAA
Glu Ile Arg Phe Phe Glu Ala Thr Thr Glu Thr Thr Leu Val Trp Phe Thr Thr Thr Thr Thr Thr
650
CGC CAC CCG ATG CCG TTT GAA CCG ATG CCG GTT TCC CTT TCT CCG TCC CCG AAC GTC CCG
Arg His Cys Met Cys Phe Glu Cys Met Arg Val Ser Val Ser Thr Ser Thr Thr Thr Thr
700
GAC TAC CCG ATC GAA AAA CCG ATG GAA TTT CCG CCG CCG GTC ATC ACT CCG TCA GAC TCC
Glu Thr Thr Thr Thr Thr Lys Ala Met Glu Phe Cys Thr Thr Arg Val Thr Thr Thr Thr Thr
750
ARG CCG ACT GTA GTT GAT GAA AAC GGA TTC CCG AAA GAG AAA CTG GCA CCG GTT ATC GAA
Ser Cys Thr Val Val Asp Ala Glu Ser Cys Phe Thr Lys Glu Lys Leu Ala Arg Leu Thr Thr
800
ATC AAA CCG ACC CCG GAT GAT CCG GTC GCA GAT TAC CCG AAA GAA TTT CCG GTC CCG TAT
Ile Lys Ala Ser Arg Asn Glu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
850
CTC GAA CCG GAA CAG CCG TGG TCT GYA CCG GTT ATC CCG CCG CTG CCG TCC CCG ACC CAG
Glu Ala Glu Thr Glu Glu Phe Trp Ser Leu Phe Val Asp Ile Thr Leu Phe Cys Ala Thr Thr Glu
900
AAG CTC GAT GAT GTT GAG CCG CCG GAT GAA GTT ATC ACT GAT CCG AAT CCG GTT AAA TCC CCG
Asn Thr Leu Asp Val Asp Ala Thr Ala His Glu Leu Thr Thr Asn Cys Val Lys Ala Val Thr
950
GAA CCG GCA GAT ATG CCG CCG ACC ATC CAG CCG ACT GAA CCG GTC CAG CAG CCG CCG GTA
Glu Cys Ala Asn Met CCG CCG Thr Thr Thr Glu Ala Thr Thr Glu Leu Phe Cys Thr Thr Thr Thr
1000
CTA TTT GCA CCG GCT AAA GCG CCG GAT CCG GCG CCG CCG CCG GTC CCG GCA TCG CCG GAA ATG
Leu Phe Ala Phe Phe Cys Lys Ala Thr Asn Thr Glu Thr Thr Thr Thr Thr Thr Thr Thr Thr
1050
GCA GAA AAC CCG CCG CCG CTC GCG TCG AAA CCG GAG AAA CTT GAG GCA CCG TTT CAG CAC
Ala Thr Asn Thr Ala Thr Arg Leu Glu Thr Lys Thr Lys Thr Val Asp Thr Arg Leu His His
1100
ACC ATG CTG GAT ATC CAC CAT CCG TCT GTT GAG CAT CCG GCT GAA GAT GAG GAA ACC AAC
Thr Met Leu Asp Ile His His Ala Cys Val Glu His Glu Thr Glu Thr Thr Thr Thr Thr Thr
1150
TAT GTC CAG CCG CCG AAC ATT CCG CCG TTT GTC AAG CTT CCG GAT CCG ATA CCG CCG CAG
Tyr Val Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
1200
CGT GTC ATT TAA
          CTTGTAAGTGT CCGTATGGCC CTACCGTAT CAGCGCTAAG CATGGCGACA
1250
ATTACATGCA GTTACGGCTCT AATGTAGGCC GCGCAAGCCG ACCCGCCCGC GCAAPATTTC AGCGCTTTAT
1300
GAGTATTTAA CCGATGATCC TCGCCAGGGA ACAATTCTTA TCGGCCACCG GCAATTTCTTA CTGATGCTT
1350
CCGAAACTCT CTTGTCTGTA CGATACAGCC CTTCAGTCTT ACATCGCTCT AACTTCTCTC CGACTCAGCC
1400
ACCGAGCACT ATCCCAATTA AGAATGCGCA TTTGCACTATT TTTTGTGATA TCTTCTTCCG TTTGCAACAA
1450
CTCGTGGCCG TTTGCAAGAG CAGCATTAT ATATGCCAC GCGACTTCTT CTTCAGTCTT CCGCTTTTGA

```

Fig. 2. Nucleotide sequence of the *gdhA* gene from *E. coli* K-12. Numbering of residues begins with the first A of the initial ATG methionine codon. A possible ribosome binding site at positions -15 to -11 is underlined.

Table 1

Codon usage frequency in the *gdhA* gene from *E. coli* K-12

Arg	CGC - 12	Ala	GCC - 16	Val	GTG - 9	Glu	GAA - 23
	CGT - 9		GCA - 10		GTC - 8		GAG - 8
	CGA - 1		GTC - 8		GTT - 16	His	CAT - 6
	CGG - 0		GCG - 13		GTA - 4		CAC - 4
	AGG - 0	Gly	GGC - 22	Ile	ATC - 12	Lys	AAA - 18
	AGA - 0		GGT - 21		ATT - 4		AAG - 4
Leu	CTG - 18		GGG - 4		ATA - 1	Asn	AAC - 9
	CTT - 6		GGA - 2		AAT - 8		AAT - 8
	CTA - 3	Pro	CCG - 12	Asp	GAT - 14	Phe	TTC - 11
	TTA - 1		CCC - 0		GAC - 5		TTT - 11
	CTC - 6		CCT - 2	Cys	TGC - 3	Tyr	TAT - 5
	TTG - 1		CCA - 1		TGT - 1		TAC - 5
Ser	AGC - 6	Thr	ACC - 11	Gln	CAG - 15	Met	ATG - 18
	TCA - 5		ACT - 6		CAA - 8		TGT - 5
	TCT - 4		ACA - 4				
	TCC - 5		ACG - 1				
	AGT - 1						
	TCG - 1						

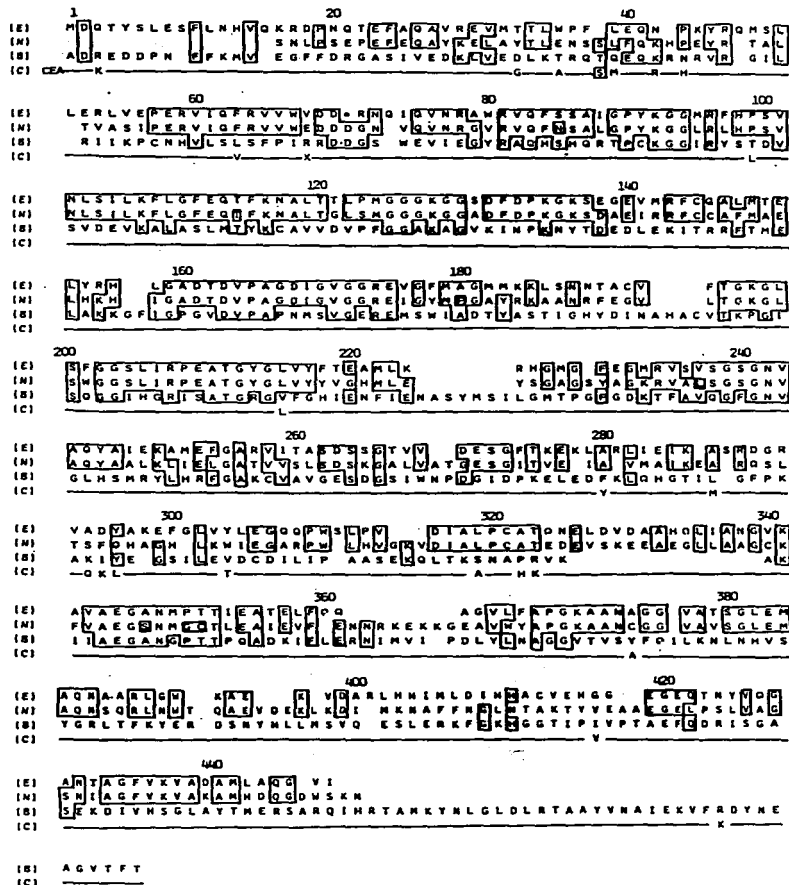


Fig. 3. Comparison of amino acid sequences of the GDH monomer from *E. coli* (E), *N. crassa* (N), bovine (B) and chicken liver (C). Numbering refers to the *E. coli* sequence. Data for sources other than *E. coli* are from Blumenthal et al. (1975). Identical amino acid sequences are boxed. The sequence of the chicken enzyme (solid line) is identical to that of the bovine, except at the positions where specific substitutions are shown.

flanqueadoras 5' y 3'. Debajo de la secuencia nucleotídica de la región estructural se presenta la secuencia de aminoácidos deducida a partir de la secuencia de nucleótidos.

El análisis de la secuencia nucleotídica nos permitió encontrar un frente de lectura abierto que inicia con el triplete ATG (metionina) y termina con el triplete TAA. Este frente de lectura de 1441 pb, codifica para 447 aminoácidos. El peso molecular calculado para el monómero de la enzima a partir de la secuencia, es de 48,589.5 y está de acuerdo con los valores reportados previamente basados en movilidades electroforéticas relativas (Sakamoto et al., 1975).

b).- Frecuencia de uso de codones.

En la tabla I de este mismo artículo, se presentan los datos de la frecuencia de uso de codones. Como se puede ver, hay un cierto uso preferencial de codones, que es similar al de aquellos genes cuyos ARN mensajeros son traducidos eficientemente (Grosjean y Fiers, 1982; Gouy y Gautier, 1982).

c).- Comparación entre las secuencias de aminoácidos de diversas GDHs.

En la figura 3, del artículo en que se reportó la secuencia nucleotídica del gene gdhA, se presenta una comparación entre la secuencia de aminoácidos de la GDH de Escherichia coli deducida a partir de la secuencia del gene, y las partes correspondientes de otras secuencias de las GDHs de otros organismos. Esta figura, muestra que los monómeros de las GDHs NADP-dependientes de E. coli y N. crassa tienen aproximadamente el mismo número de residuos (450), mientras que los dos monómeros NAD-dependientes están compuestos de 501 residuos.

Aun más, los monómeros de las GDHs de E. coli y N. crassa, son más homólogos entre sí que con los monómeros de las GDHs NAD-dependientes.

Región flanqueadora 3' del gene gdhA.

a).- Características estructurales.

Un análisis por computadora, en el que se hizo una búsqueda de secuencias con simetría de diada en una región de aproximadamente 200 pb, posterior al triplete de terminación, reveló que un ARNm que porte esta región, puede formar diferentes estructuras estables de tallo-asa. Una de ellas, una estructura estable de tallo-asa, la cual empieza en el nucleótido 87 después del codón de terminación (UAA), tiene las características generales de un terminador rho-independiente (Adhya y Gottesman, 1978; Holmes et al., 1983). Un esquema de este posible terminador se presenta en la figura 5 del presente trabajo.

Las otras secuencias palindrómicas relevantes que se encuentran en la región flanqueadora 3' del gene gdhA, se presentan en la figura 1 del artículo que se anexa, referente a las secuencias REP del gene (Becerril et al., 1985). Estas secuencias, pueden formar una estructura parecida a un ARNt. Un esquema de esta última estructura, se presenta en la figura 2 de la publicación. Es importante recalcar que el 35% de sus nucleótidos se encuentran ubicados en posiciones homólogas a las del ARNt para glutamato de E. coli (Sprinzi y Gauss, 1983), y que incluye a las dos secuencias REP opuestas. Estas dos secuencias, empezando en las posiciones 1353 y 1420 (ver figura 1, Becerril et al., 1985), son marcadamente homólogas a las secuencias REP consenso reportadas por Stern et al. (1984), y como se puede

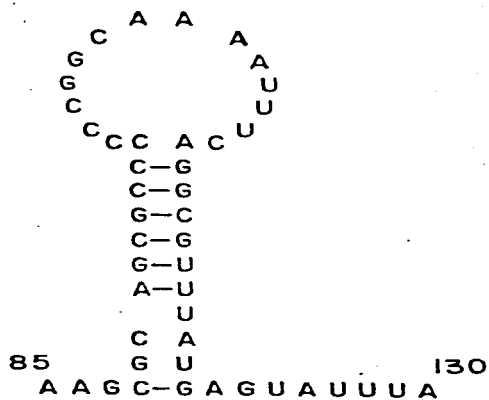


Figura 5. Esquema de un posible terminador rho-independiente aportado por la secuencia nucleotídica que se inicia en la posición 85 a partir del codón de terminación del gene gdhA.

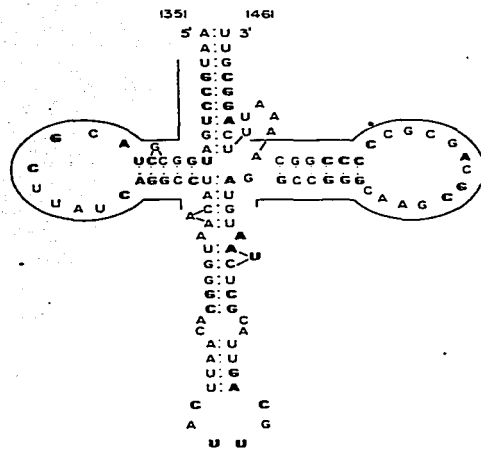


Fig. 2. A tRNA-like structure resulting from folding of the RNA transcript carrying the first 120 nt after TAA stop codon (for numbering see Fig. 1). The nt in common with glutamate tRNA appear in bold characters. The location of REP sequences are indicated by solid lines.

apreciar en la figura 1, ellas pueden ser estructuradas al menos en dos formas alternativas.

b). Construcción de mutantes por delección de REP.

Stern et al. (1984), han propuesto que las secuencias REP podrían estar involucradas en la regulación de la expresión genética de los procariotes. En un esfuerzo por entender el posible papel de esta clase de secuencias en la regulación de la concentración de la GDH en la célula bacteriana, decidimos eliminar una de estas secuencias REP localizadas en la región flanqueadora 3' del gene gdhA. Este experimento se realizó tomando ventaja de la presencia de dos sitios únicos de StuI localizados en esta región flanqueadora, fuera de la región codificadora del gene gdhA (posiciones 1375 y 2375, Fig. 1), y un sitio único de NruI en el gene de resistencia a tetraciclina del vehículo de clonación pBR322 (Bolivar et al., 1977). Usando esta estrategia, se derivaron los plásmidos pBB5 y pBB6 a partir del plásmido pSAE4. Estos dos derivados fueron construidos con el objeto de poner diferentes secuencias de ADN hacia abajo del gene gdhA y de esta manera evitar una interpretación incorrecta de los resultados. Se compararon las actividades de la GDH sintetizada a partir de estos tres plásmidos, usando como hospedero la cepa PA340. Las concentraciones finales de la GDH, fueron reducidas al menos un 30% (generalmente 50%), por efecto de las delecciones, al compararlas con la de células que portan el plásmido pSAE4 (Fig. 3, publicación REPs de gdhA, Becerril et al., 1985). Es interesante puntualizar que este efecto es independiente del medio de crecimiento. Estos resultados están de acuerdo con los resultados reportados por Stern et al. (1984), quienes demostraron que al eliminar las secuencias REP en

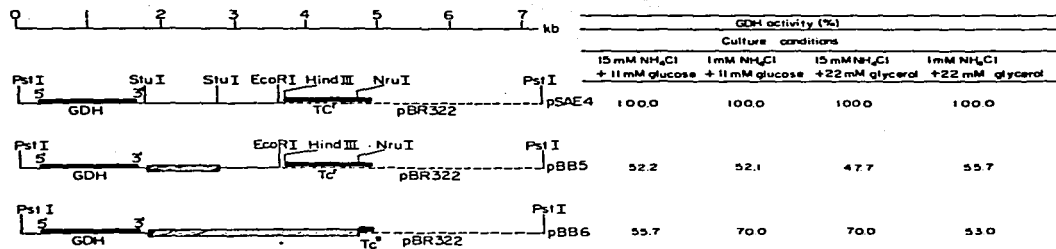


Fig. 3. Physical maps of the plasmids pSAE4, pBB5 and pBB6 linearized at the *Pst*I site. The dashed lines represent the vector pBR322, the solid lines represent *E. coli* chromosomal inserts and the thick bars indicate the locations of the *gdhA* and *Tc^R* genes. The hatched areas indicate deleted 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.

el operón hisJ-P de Salmonella typhimurium, disminuye la concentración de la enzima codificada por el gene localizado hacia arriba de las secuencias REP.

c).- Análisis detallado del banco de secuencias (GenBank).

Estos resultados nos llevaron a implementar un estudio más detallado de las secuencias REP y sus secuencias aledañas. Una búsqueda por computadora en el banco de datos GenBank (septiembre, 1984), se realizó usando el segmento más conservado de la secuencia REP:

C

5'-CTTATC GGCCTA-3' y su secuencia complementaria, en lugar de la

A

usada por Stern et al. (1984). Estos investigadores reportaron una serie de secuencias REP de las cuales se derivó un consenso; nuestros resultados concuerdan con los suyos. Sin embargo, nos gustaría enfatizar ciertos aspectos novedosos. En aquellas unidades transcripcionales en las que REP ha sido localizada, y se sabe donde termina la transcripción (9 de 25), REP está localizada dentro del transcrito. Con base en este criterio, decidimos comparar las secuencias REP, tomando en consideración su orientación con respecto al gene identificado más próximo (gene proximal), en relación a la dirección de transcripción. Mediante tal comparación fuimos capaces de agregar una base extra al consenso propuesto y también pudimos notar ciertos aspectos de la región central (no apareada) de las secuencias REP. Estos resultados se presentan en las Tablas I, II y III de la publicación antes referida. La Tabla I muestra las unidades transcripcionales bacterianas en las cuales se han encontrado las

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 Higgings 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*, *malE*, *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
aroG		49	—
mtlA		41	—
folA		24	—
lacY		9	—
* dhvA (S.t.)		0	—
* rpoD		3	—
* uvrD		b	—
MI RNA		c	3,5
spot 42		60	25
hisJ (S.t.)		47	20
hisG (S.t.)		6	25
glyA		36	35
* supBE tRNA		d	5
* tyr tRNA		e	13
gdhA		9	3,4
* glnALG		59	1,4
plsB		12	3,6
lamB		9	7,14
ndhA		12	1,4
trpR		4	15,26
glnS		20	30,35
sdhA		NA	NA
sucB		NA	NA
* araBAD (S.t.)		14 (f)	6,11

^a *dhvA* (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 polipeptide 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 *araC* 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).

secuencias REP. El número de copias de las secuencias REP, la orientación con respecto al gene proximal la distancia desde el codón de terminación de este gene y la distancia entre las secuencias REP, están indicadas en esta Tabla. Es importante hacer notar que algunas de estas secuencias REP sobrelapan con promotores propuestos, o están localizadas al final de operones que codifican para ARNts. Además, nos gustaría puntualizar que algunas de las secuencias REP nuevas aquí presentadas, no son del todo complementarias.

Esta nueva búsqueda nos permitió identificar 16 nuevas secuencias REP, localizadas en 9 unidades transcripcionales (Tablas II y III). La secuencia REP consenso presentada en la Tabla II, es la propuesta por Stern et al. (1984), una base adicional (T), en el extremo 5'; la secuencia REP consenso complementaria presentada en la Tabla III, incluye una base adicional (A) localizada en el extremo 3'. La figura 4 muestra dos posibilidades alternativas en la que estas secuencias REP se pueden estructurar.

Otro aspecto interesante que se derivó de esta búsqueda es la existencia de secuencias tipo REP formando parte de genes que codifican para ARNs estables (Tabla IV).

Cuando se analizaron las regiones de ADN en la vecindad de las secuencias REP, notamos que en las secuencias de las unidades transcripcionales de glyA, plsB y MIARN, la secuencia: 5'-GCAAATTCAATATATTGCA-3', se encuentra conservada. Aun más, se han encontrado secuencias similares en gdhA y glnS y entre otras unidades transcripcionales (no mostradas). Finalmente, buscamos esta secuencia en el banco de secuencias, y la encontramos en algunos sitios incluyendo la región espaciadora entre el ARN 58 y el ARNt de valina

TABLE II
Homology between REP sequences

Name	Sequence
Consensus	T G C C G G A T G C G G C G T 0 to 6 nt G A C G C C T T A T C C G G C C T A C T A G A A C G T A
<i>malE</i>	G C C C G G A T G C C G G T G A C A C C G C T T T G A T T C C G G C C T A C C
<i>sucB</i> (2)	G C C C T G G A T G T G G A C C C G G T T G T T A A A C G G T G G C C C C C T T A C C C
<i>sucB</i> (4)	G C C C T G G A T G T G G A C C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>sdhA</i> (2)	G C C C G G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
M1 RNA	T G C C C C T G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>hisG</i>	T G C C C C T G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>polA</i>	T G C C C C T G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>glyA</i> (1)	T G C C C C T G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>qdhA</i>	T G C C C C T G G A T G C C G G G C C G G T T G A G A A C C G G C C C C C T T A C C C
<i>qInALG</i>	T G C C C C T G G A T G C C G G G A C C C G G C T A A A T T G A G A A C C G G C C C C C T T A C C C
<i>trpR</i> (2)	T G C C C C T G G A T G C C G G G A C C C G G C T A A A T T G A G A A C C G G C C C C C T T A C C C
<i>hisJ</i> (S.t.)	T G C C C C T G G A T G C C G G G A C C C G G C T A A A T T G A G A A C C G G C C C C C T T A C C C
<i>lamB</i> (2)	T G C C C C T G G A T G C C G G G A C C C G G C T T T G C G C C G C C C C C C C C T T A C C C
<i>glnS</i> (3)	T G C C C C T G G A T G C C G G G A C C C G G C T A A A T T G A G A A C C G G C C C C C T T A C C C
<i>plsB</i>	T G C C C C T G G A T G C C G G G A C C C G G C T A A A T T G A G A A C C G G C C C C C T T A C C C
<i>mtiA</i>	T G C C C C T G G A T G C C G G G A C C C G G C T C G G T T C G C C C G C C C C C C C T T A C C C
<i>foiA</i>	C G C C C G G A T G C C G G G C C G G C C T C G G T T C G C C C G C C C C C C C T T A C C C
<i>lacy</i>	T G T T C G G A T G C C G G G C C G G C C T G T A C C G T C C T T A T T C C C G G C C C C C T T A C C C
<i>uvrD</i>	T G C C C C G G A T G C C G G G C C A C T C G T A C C G T C C T T A T T C C C G G C C C C C T T A C C C
<i>tyr tRNA</i>	T G C C C C G G A T G C C G G G C C A C T C G T A C C G T C C T T A T T C C C G G C C C C C T T A C C C
<i>ndhA</i> (2)	T G C C C C T G G A T G C C G G G C C A C T C G T A C C G T C C T T A T T C C C G G C C C C C T T A C C C
<i>ndhA</i> (3)	T G C C C C T G G A T G C C G G G C C A C T C G T A C C G T C C T T A T T C C C G G C C C C C T T A C C C
<i>supBE tRNA</i>	T G T T C G G A A T G C C G G A C A G G T T G A C A C C G T C C T T A T T C C C G G A C C C T A C C
<i>supBE tRNA</i>	T G T T C G G A A T G C C G G A C A G G T T G C C G C C G G T T T A C C C G A C C C T A C C
<i>araBAD</i> (1) (S.t.)	T G C C C G G A T G C C G G G A C C T G G C A C C G T T T A C C G G G C C C T A C 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 *malE*, *hisG*, *sucB* and *sdhA*, which are from Stern et al. (1984).

TABLE III
Homology between REP sequences

Name	Sequence
Consensus	G T A G G C C G G A T A A G G C A T T A T G C G T C C A T C A G G C A T G A T A A G G C A T T A T G C G T C C A T C A G G C A 0 to 4 nt A C G C G C A T C A G G C A
sucB (1)	G T A G G A C C G G A T A A G G C A T T A T G C G T C C A T C C G G C C
sucB (3)	G T A G G G T C C C G G A T A A G G C G C G C G C A A G C C C C G G C C A T T C C A G G C C
sdhA (1)	G T A G G C C T G A T A A G G C G C G C G C A A G C C G T C C G C C A T T C C A G G G C
sdhA (3)	G T A G G C C C G G G G C A A G G G C C G C G C A A G C C G T C C G C C A T T C C A G G G C
hisJ	G T A G G C C C G G G G C A A G G G C C G C G C A C C A A G C G T C C G C C A T T C C A G G G G C G
aroG	A C G G G C C C G G G A T A A G G T C G G C G C G C A A A G C G T C C G C C A T T C C A G G G C A
poiA	G T A G G G C C G G G A T A A G G C G C G C G C A A G C C G T C C G C C A T T C C A G G G C C A
glyA	G T A G G G C C G G G A T A A G G C G C G C G C A A G C C G T C C G C C A T T C C A G G G C C A
gdhA	G T A G G C C C G G G C A A G G C G C G C G C A A G C C G T C C G C C A T T C C A G G G C A
glnALG	G T A G G G C C G G G G A G C A A G G G C G T G A G T G G C C C C G T C C A A C C C G T G A A
trpR (1)	G T A G G G C C C T G G A T A A G G A C C G C T G G C A T T A G C C G C C G G C C A T T C C A G G G C A
trpR (3)	G T A G G G C C A T G G A T A A G G A C C G C T G G C A T T A G C C G C C G G C C A T T C C A G G G C A
lamB (1)	C T T G G G C A T A A G G G C C G T T T T A C C G G C C C G G C C A T T C C A G G G C C A
lamB (3)	G T A G G C C C G G G A T A A G G G C C G T T T T A C C G G C C C G G C C A T T C C A G G G C A
glnS (1)	C C T A T G G C C G G G A T T G G G G C C G T T T T A C C G G C C C G G C C A T T C C A G G G C A
glnS (3)	G T A G G G C C C G G G A T T A A G G G C C G T T T T A C C G G C C C G G C C A T T C C A G G G C A
dhuA (S.t.)	G T A G G G C C C T G G A T T A A T G G G C C A C G A T A A C C G G G C C C A T T C C A G G G C A
plsB	G T A G G G C C C T G G A T A A G G G C C G T T T T A C C G G C C C G G C C A T T C C A G G G C A
ndhA (1)	G T A G G G T C C G G G A T A A G G C C G T A C C A C C G G C C G C C A T T C C A G G G C A
tyr tRNA	G T A G G G T C C G G G A T A A G G C C G T A A A A C G G C C G C C A T T C C A G G G C A
M1 RNA	G T A G G G C C C T G G A T A A G G G C C G T A A A A C C G G C C G C C A T T C C A G G G C A
hisG	G T A G G G C C C G G A T A A G G G C C A G T T T A T G T C G C C A T T C C G G C A
araBAD (2) (S.t.)	G T A G G G C C C G G A T A A G G G C C A G T T T A A G C C C A T T C C G G C A
araBAD (3) (S.t.)	G T A G G G C C C G G A T A A G G G C C A G T T T A A G C C C A T T C C G G C A

Nucleotide sequences of the complementary consensus REP sequences and of the transcriptional units shown in Table I. The symbols are as in Table II.

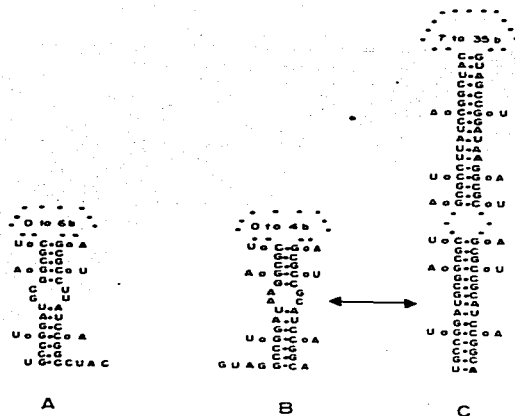


Fig. 4. Stem and loop structures that can be potentially formed within the REP consensus sequences. Structures A and B, were generated from consensus sequences presented in Tables II and III, respectively. Structure C was generated by pairing structures A and B. Note that the paired regions of A and B are identical; conserved sequences CUAC and GUAG, found at the 3' end in A and at the 5' end in B, and also the central bubble on each structure, confer upon them directionality with respect to the proximal gene. It is important to point out that the REP consensus sequence proposed by Stern et al. (1984), is the same as the REP consensus sequence in this figure (A and B), except that there is an additional uridine at the 5' end of A and an additional adenine at the 3' end of B.

TABLE IV

REP-like sequences found as part of genes that code for stable RNA molecules

Name	Sequence
Consensus	G T A G G C C T ^G G A T A A G ^G C G T 0 to 4' ^m A C G T C G C A T C ^C A G G C A
B. subtilis tRNA ile	G c A c G C C T G A T A A G C G T g A C G T C G g T g g t t CgA (24)
E. coli tRNA ile	G c A G G C C T G A T A A G G G T A g G T C G g T g g t t C A (24)
E. coli 16S RNA	G T A G t C C G G A T t G G a G T ctgca A C G T C G AcT C C a t g A (25)
E. coli 23S RNA	G T A a G C C T G c g A A G G C G T gt G CtG T GaggCA T C CtG G A (28)

At the right margin, in parentheses, are specified the numbers of bases (in capital letters), out of 33, that are homologous to the consensus REP sequence. Non homologous bases are in lower-case letters.

de Bacillus subtilis (Green y Vold, 1983).

DISCUSION

Nuestro laboratorio ha reportado la clonación (Sánchez-Pescador et al, 1982), la secuencia de la región regulatoria y la región del gene que codifica para enzima glutamato deshidrogenasa de Escherichia coli K-12 (Valle et al., 1983; 1984). También se ha propuesto la existencia de tres posibles promotores, un posible sitio de unión a CRP y un probable sitio de unión al ribosoma (Valle et al, 1983).

Con el propósito de caracterizar tanto estructuralmente como a nivel de regulación este gene, se determinó su secuencia nucleotídica. La secuencia de aminoácidos deducida a partir de la nucleotídica, reveló una proteína constituida por 447 residuos de aminoácido, con un peso molecular de 48589.5 Daltones, lo cual está de acuerdo con los datos de peso molecular reportados para esta proteína (Sakamoto et al, 1975).

El análisis de la frecuencia de uso de codones muestra que se usan preferencialmente ciertos tripletes y otros son evitados. En este sentido es notorio que los tripletes CGA, AGA, CUA, GGG, GGA, AUA, son prácticamente excluidos. Es importante hacer notar que estos tripletes han sido propuestos como moduladores de la traducción de algunos genes y que los ARN mensajeros de genes, que por su función es vital que sean traducidos rápidamente, evitan el uso de dichos tripletes (GrosJean y Fiers, 1982; Gouy y Gautier, 1982).

El análisis global de la estrategia de codificación de este gene sugiere que el ARNm para la GDH de E. coli, se traduce con

una alta eficiencia y que se comporta muy similarmente a los de proteínas muy abundantes en la célula. Todo lo anterior podría sugerir la existencia de una presión de selección hacia este gene, aunque esto parece ser contrario a lo esperado, ya que se ha demostrado que cepas de E. coli gdh- no presentan un fenotipo claro y que parece ser un gene dispensable (Tyler, 1978).

La secuencia de aminoácidos del monómero de la GDH de E. coli se comparó con las secuencias de otras GDHs. Con el fin de obtener máxima homología, fue necesario introducir espacios en las secuencias. Con los límites impuestos por este alineamiento, podemos observar una excelente homología entre las secuencias de E. coli y N. crassa. El examen de estos datos indica que muchos de los residuos conservados probablemente tienen una importancia conformacional más que catalítica como ha sido sugerido por Smith et al (1975). y Blumenthal et al (1975). Por ejemplo la mayoría de los 37 residuos de glicina conservados entre N. crassa y E. coli deben ser importantes para la conformación, ya que se ha demostrado que los residuos de glicina se encuentran generalmente en porciones no helicoidales de las proteínas.

La mayor homología entre estas cuatro proteínas se encuentra en la primera mitad del monómero. Esta sección de la proteína incluye un residuo de lisina altamente reactiva (Blumenthal et al, 1975) en la posición 128 de la secuencia de E. coli. También incluye probablemente la región involucrada en la unión del acetoglutarato y en la actividad catalítica. Estas suposiciones están apoyadas por estudios de modificación química y mutantes (Brett et al, 1976; Julliard y Smith, 1979; Haberland y Smith, 1980; Austen et al,

1980; Ashby et al, 1981).

Además de las identidades observadas, un gran número de los substitutos presentan el mismo carácter de hidrofobicidad y coeficiente de accesibilidad (Cid et al, 1982). Estas observaciones sugieren que estas proteínas pueden ser mas homólogas entre sí a nivel tridimensional que a nivel de su secuencia de aminoácidos. En este sentido y utilizando el método de Cid et al, (1982) para la predicción de estructura secundaria de proteínas, hemos encontrado que en algunas regiones donde la secuencia de aminoácidos entre E. coli y bovino no se conserva (pero si se conserva entre N. crassa y E. coli), ambas proteínas pueden formar la misma estructura secundaria (datos no mostrados).

Región de regulación 5'.

El análisis de la secuencia (aproximadamente 200 pb) de la región inmediatamente anterior al codón de iniciación del gene (Fig. 1 de esta Tesis), ha permitido proponer la existencia de tres posibles promotores por criterios de secuencia consenso (Valle et al, 1983). Estos supuestos promotores han sido denominados P1, P2 y P3, en relación al orden que guardan con respecto a la metionina inicial, siendo P1 el mas cercano.

Estudios de transcripción in vitro, indican que el fragmento TaqI- HpaII (Fig.1), une ARN polimerasa. El promotor presente en este caso sería P1 (Valle et al, 1983).

Por otro lado, en experimentos in vivo, el fragmento HpaII-HpaII, clonado en el sitio de ClaI del plasmido pBR327, ensayado en minicélulas da un péptido de 63 aminoácidos independientemente de

la orientación en la que se haya clonado (Valle et al , 1983). Estos datos indican que este fragmento contiene un promotor propio. Sin embargo, esto no sucede con el fragmento TaqI-HpaII clonado y ensayado de la misma manera, es decir, no se produce ningún péptido.

Estudios preliminares de mapeo por nucleasa S1, indican que bajo condiciones de crecimiento en alta concentración de amonio, la transcripción es iniciada enfrente del supuesto promotor 3 (Valle et al, 1983).

Los datos anteriores sugieren que la región comprendida entre HpaII y TaqI (Fig.1), es esencial para la expresión del gene gdhA.

Otros aspectos importantes de la región regulatoria involucran a una serie de secuencias repetidas directas, las cuales podrían ser sitios de interacción con proteínas regulatorias y/o conformar dominios estructurales que le confirieran una arquitectura especial al ADN y que esta tuviera alguna influencia sobre la regulación de la expresión del gene. Recientemente, se ha reportado la participación de una serie de 5 secuencias repetidas directas localizadas en la región inmediatamente hacia arriba del promotor 2 del operón glnALG de Salmonella typhimurium, en la conformación de igual número de sitios de interacción con el producto polipeptídico del gene glnB (ntrC) (Hirschman et al, 1985).

Otra característica importante de la región regulatoria es que tomando como referencia al sitio de TaqI (Fig.1), hacia arriba de este sitio, se encuentra una región rica en Ts y hacia abajo de dicho sitio, una rica en As. Esta característica podría influir en la

arquitectura del ADN.

Se ha propuesto que la secuencia C-A(5-6)-T presente con una cierta periodicidad induce un combamiento o doblamiento en el ADN (Marini et al, 1982; Wu y Crothers, 1984). Variantes de esta secuencia se encuentran hacia abajo del sitio de TaqI, sugiriendo que esta región podría tener algún tipo de combamiento y que de alguna manera afectara la expresión del gene gdhA.

Identificación del promotor.

Los resultados de los experimentos de extensión de primero, sugieren fuertemente que al menos en condiciones de crecimiento en glicerol o glucosa como fuente de carbono y 1 mM o 15 mM NH_4Cl como fuente de nitrógeno, el promotor 3 es funcional tanto en cromosoma como en plásmido (CS520 comparada con PA340/pSAE4). Por un lado, en todas las condiciones antes mencionadas se detecta el mismo sitio de inicio de la transcripción (banda de ADNC de 117 nucleótidos, Fig. 3), tanto en la cepa silvestre (CS520), como en la cepa con plásmido (PA340/pSAE4). Por otro lado, las manipulaciones hechas en la región regulatoria, es decir, la inserción de un par de bases en TaqI, uno y dos adaptadores sintéticos de EcoRI en el nuevo sitio de NruI, en las cuales el inicio de la transcripción se ve modificado por la incorporación de distintas secuencias interpuestas entre el -10 del promotor 3 y el inicio de la transcripción original, permiten afirmar que el promotor 3 es el único funcional, al menos en las condiciones ensayadas.

Esta afirmación se deriva tanto del hecho de que independientemente de las condiciones de crecimiento ensayadas, la

transcripción se inicia en el mismo sitio (frente al promotor 3), ya sea en plásmido o en cromosoma. Así mismo, independientemente de las manipulaciones realizadas, la transcripción siempre se inicia dentro de los 5-8 nucleótidos hacia abajo de la T en la posición -8 de este promotor. Esta observación está de acuerdo con los datos reportados tanto in vitro como in vivo para un número importante de genes de E. coli (Hawley y McClure, 1983), es decir, el sitio de inicio de la transcripción se encuentra generalmente dentro de los 5-8 nucleótidos hacia abajo de la T en la posición -8 de la caja de Pribnow.

Finalmente, el hecho de que al eliminar la región que comprende a este promotor (EcoRII-TaqI), no se detecte ningún transcrito, confirma nuestra afirmación.

Un punto más que nos gustaría comentar con respecto a la región regulatoria es la observación de que la inserción de varios "linkers" de EcoRI en el sitio de EcoRII, no modifica el sitio de inicio de la transcripción con respecto a la manipulación hecha en el plásmido padre (pSAE432 comparado con pSAE442; Fig. 4). Este resultado sugiere que la región del sitio de EcoRII, no está involucrada de manera importante en el inicio de la transcripción a partir del promotor 3. Sin embargo, es importante hacer notar que la intensidad de la banda es menor que lo que se esperaba. Esto podría indicar que se está teniendo un efecto parcial sobre la transcripción del gene gdhA a partir del promotor 3.

Para continuar con el estudio de este posible efecto, actualmente se está manipulando el plásmido que lleva el gene completo (pSAE4), con el objeto de reproducir la manipulación, es decir, la inserción de varios "linkers" de EcoRI en el sitio de EcoRII de la

región regulatoria del gene, y determinar el nivel de transcrito específico, la actividad específica de GDH, el número de copias del plásmido, para saber a que nivel se estaría afectando la expresión del gene, si es que realmente se está afectando de manera importante.

El último punto a considerar referente a la región regulatoria es el relativo al posible sitio de unión a CRP, sobrelapado con el -35 del promotor 3 (Fig. 1). La participación de CRP en la regulación positiva del gene *gdhA* de *E. coli* está reportada en la literatura (Prussiner *et al*, 1972).

Dado que la reacción catalizada por la GDH, involucra tanto al metabolismo de nitrógeno como al de carbohidratos sería de esperar que el gene que codifica para esta enzima estuviera regulado por CRP.

Ya ha sido discutido que al menos en las condiciones de crecimiento ensayadas en el presente trabajo, todo parece indicar que el promotor P3 es el único funcional, al menos en las condiciones metabólicas probadas. La ubicación del sitio de CRP sobrelapado con el -35 de este promotor indica que al menos en esas condiciones metabólicas, la activación por CRP no es clara, en función de que no se han detectado transcritos que inicien en P2 o P1. Sin embargo, no se descarta la posibilidad de que en otras condiciones metabólicas P2 o P1 puedan ser utilizados.

Si P3 fuera el único promotor funcional, el efecto activador de CRP se podría explicar en función de que se ha demostrado que el complejo CRP-AMPC, puede funcionar como activador al ejercer una función de antiterminador (Guidi *et al*, 1984). Una explicación

alternativa para el efecto activador de CRP, es que se pueda pegar al mismo sitio que la ARN polimerasa pero en la cara contraria del ADN. En este sentido, ha sido demostrado que el producto polipeptídico del gene CII del fago lambda activa ciertos promotores de éste, cuando se une en la otra cara del ADN con respecto a la ARN polimerasa (Ho et al, 1983). Aun más, existe al menos otro caso en el que un sitio de CRP sobrelapa el -35 del promotor y tiene un papel de activador (Musso et al, 1977).

Region flanqueadora 3'.

El análisis de aproximadamente 800 pb después del triplete de terminación del gene estructural gdhA, no reveló ningún frente amplio de lectura que permitiera la codificación de otro polipeptido, lo cual sugiere que este gene no forma parte de un operón. Basado en este supuesto, es de esperarse la existencia de un terminador de la transcripción.

El análisis de la secuencia de los primeros 200 pb de esta región flanqueadora 3' del gene gdhA, revela que un transcrito que porte esta secuencia, tendría la capacidad potencial de formar diversas estructuras secundarias. Una de ellas, una estructura estable de tallo y asa, empezando a 87 bases después del codón de terminación (Fig.5), tiene las características de un terminador rho independiente (Adhya y Gottesman, 1978; Holmes et al, 1983).

Otras estructuras alternativas son las siguientes: Empezando a 8 nucleótidos después del codón de terminación, se encuentra una secuencia marcadamente homóloga (28 de 29 identidades), con la secuencia consenso propuesta por Higgins et al., (1982), como

un elemento regulatorio novedoso de los genes procariotes. Esta secuencia también tiene la capacidad potencial de formar una estructura de tallo-asa en el ARNm.

Recientemente, Gilson et al., (1984), han reportado una familia de secuencias palindrómicas dispersas en el genoma de E. coli y S. typhimurium. Estas secuencias llamadas unidades palindrómicas, tienen una secuencia consenso similar a la del elemento regulatorio "novedoso" de Higgins et al., (1982).

Por otro lado, Stern et al., (1984), han publicado una secuencia consenso ligeramente diferente para ese elemento, la cual ha sido denominada secuencia REP (repetitive extragenic palindromic).

Dicho elemento extragenico es de alrededor de 35 nucleótidos de largo, incluye una secuencia repetida y puede ocurrir en copias adyacentes múltiples, con orientaciones generalmente opuestas entre cualquier par de ellas. Aparentemente la orientación, posición y distribución de las secuencias REP, no siguen una regla clara (Stern et al., 1984).

Con estas propiedades de las secuencias REP, se buscaron otras de ellas en el gene gdhA y se encontró una segunda secuencia REP, en sentido opuesto a la primera y separadas entre sí por 34 bases (ver Fig. 1, publicación sobre REP). Esta última secuencia REP es menos homóloga que la primera, en relación a la secuencia consenso reportada por Stern et al., (1984). Estas dos secuencias REP de gdhA, pueden arreglarse al menos en dos estructuras alternativas, cada una por separado o apareadas en una sola estructura, con un tallo de 27 pb.

Otra secuencia que destaca es una que tiene la capacidad potencial de formar una estructura semejante a un ARNt. Es notable que en esta última secuencia de 110 nucleótidos, 40 de ellos se encuentran en posiciones homólogas a las del ARNt de glutámico de E. coli (Fig.2)(Sprinzi et al, 1983).

Stern et al., (1984), han propuesto que las secuencias REP, podrían estar involucradas en la regulación de la expresión genética en los procarlotes.

Con el objeto de entender el papel de las secuencias REP en la determinación de la concentración final de la GDH en la célula bacteriana, se decidió quitar la segunda secuencia REP. Este experimento se realizó tomando ventaja de la presencia de dos sitios únicos de StuI, localizados a 35 y 1035 nucleótidos posteriores al codón de terminación y un sitio único de NruI situado al final del gene de resistencia a tetraciclina del plásmido pSAE4, el cual porta el gene gdhA y sus regiones flanqueadoras 5' y 3'.

Mediante la digestión con estas enzimas fue posible quitar 1000 y 3000 pb de ADN del plásmido pSAE4. Ambos fragmentos incluyen la segunda secuencia REP.

Al determinar la actividad específica de la GDH en las células que portan los plásmidos derivados por digestión, se pudo observar un decremento de dicha actividad al menos de un 30% (generalmente 50%), independientemente de la composición del medio, número de copias del plásmido y secuencias hacia abajo de la primera secuencia REP (Fig. 3, publicación sobre REP).

Resultados similares fueron obtenidos por Stern *et al.*, (1984), cuando quitaron las dos secuencias REP del operón hisJ-P del cromosoma de S. typhimurium.

Otro dato similar se obtuvo al efectuar una eliminación de las secuencias REP del gene glyA de E. coli (Plaman y Stauffer, 1985), en donde también encuentran una disminución de la actividad específica del producto polipeptídico codificado por el gene que se encuentra hacia arriba de las secuencias REP (glyA).

Los datos anteriores podrían apoyar la idea de que en el caso de las eliminaciones de la segunda secuencia REP del gene gdhA, estarían teniendo un efecto sobre la estabilidad del mensajero de la GDH. Resultados preliminares en gdhA indican que este es el caso. Sin embargo no se descarta la posibilidad de que al mismo tiempo las secuencias REP, puedan estar cumpliendo otros papeles en la regulación del gene gdhA.

Una explicación alternativa para la disminución de la actividad específica, es que el transcrito que carece de la segunda secuencia REP, sea traducido menos eficientemente. Esta suposición obliga a pensar que las secuencias REP pudieran interaccionar con la región 5' del mensajero favoreciendo la entrada del ribosoma y que cuando se quitan ambas o solo una de ellas, esta interacción quede desfavorecida.

Si la vida media del mensajero es dependiente de la estabilidad u otra propiedad de las estructuras de tallo-asa, entonces esto abre la posibilidad de que las secuencias REP sean tan importantes como la fuerza de un promotor, para determinar la concentración final de una

proteína.

Interesantemente, la conservación de la secuencia nucleotídica de las secuencias REP tanto en E. coli como S. typhimurium, es muy pronunciada. Si cualquier estructura de tallo-asa es suficiente para estabilizar un transcrito, entonces es difícil explicar porqué se ha conservado la secuencia.

Nuestro interés por entender la importancia de la existencia de las secuencias REP, nos condujo a hacer una búsqueda de estas secuencias en el banco de secuencias (GenBank database, septiembre 1984), usando para ello la parte más conservada de la secuencia consenso reportada por Stern et al., (1984):

$$5' \text{ CTTATC} \begin{matrix} \text{C} \\ \text{A} \end{matrix} \text{GGCCTAC } 3' \text{ y su secuencia complementaria.}$$

Esta búsqueda reveló la existencia de 51 secuencias REP en 25 unidades transcripcionales de un total de \approx 400,000 pb de E. coli y S. typhimurium, analizadas.

Más información fue obtenida de este análisis. La secuencia REP consenso fue incrementada de 32 a 33 pb y también encontramos que en las secuencias REP consenso, las asas tienen un tamaño restringido 0-4 o 0-6 pb. Otra parte importante de información es que en 2 de 9 unidades transcripcionales nuevas presentadas en el trabajo sobre REP, en las cuales estas secuencias han sido localizadas, estos tipos de secuencias han sido encontradas sobrelapando regiones regulatorias, como en el caso de dhyA, donde una secuencia REP sobrelapa con el promotor de hisJ y uvrD, donde una secuencia REP esta localizada entre el -10 y el ATG inicial de la secuencia codificadora. Es también

interesante que secuencias tipo REP han sido encontradas en especies de ARNs estables tales como los ARNs ribosomales 16 y 23 S y también en el ARNt de isoleucina de E. coli y B. subtilis.

Stern et al., (1984) han propuesto algunos posibles papeles para las secuencias REP. Uno de ellos el cual ya ha sido discutido líneas arriba es el que puedan tener simplemente un papel pasivo donde las dos secuencias invertidas puedan servir como una barrera contra la actividad exonucleolítica 3' → 5' a nivel de ARNm. Alternativamente, las secuencias REP pueden afectar la expresión génica por medio de la alteración de la estabilidad del ARNm sirviendo como un sustrato específico para la ruptura endonucleolítica. Ellos argumentan que este podría ser el caso porque tales secuencias no habían sido encontradas en especies estables de ARN. Sin embargo, nuestros resultados muestran que al menos en las unidades transcripcionales que codifican para los ARNts de tirosina y SupBE, las secuencias REP están presentes y también que secuencias tipo REP, fueron encontradas en los ARN ribosomales 16 y 23S y en el ARNt de isoleucina. Creemos que la importancia de encontrar secuencias REP al final de operones que codifican para ARNs de transferencia, recae sobre el hecho de que, al menos en los casos de los operones de los ARNs de transferencia de tirosina y SupBE, las secuencias REP no podrían involucrarse fácilmente o relacionarse con la estabilidad del ARN. No obstante, el hallazgo de estas secuencias REP y tipo REP, en los operones que codifican para ARNs estables no elimina la posibilidad de que tales secuencias y/o la secuencia 5' GCAAATTCAATATATTGCA 3', localizada entre dos secuencias REP, podrían estar involucradas en el procesamiento de ciertas moléculas de ARN mensajero. Esta hipótesis está avalada por el hecho de que al menos para un caso, donde esta

secuencia ha sido localizada entre los genes que codifican para el ARN 5S y el ARNt de valina de Bacillus subtilis, se ha propuesto que esta secuencia está involucrada en la conformación de un locus de procesamiento (Gegenheimer y Apirion, 1981).

Hay otros posibles papeles para las secuencias REP. Uno de ellos es que sean requeridas para proporcionar especificidad para la unión de proteínas (Stern et al., 1984). Ya se ha indicado también, que al menos en el caso de gdhA, estas secuencias podrían estructurarse como parte de una molécula parecida a un ARNt. Esto sugeriría que la molécula podría ser reconocida por una enzima del tipo modificadora de ARNt. Datos similares han sido publicados por Ames et al., (1983), en el sentido de que una estructura parecida a un ARNt se encontró formando parte de la región líder del operón his de Salmonella typhimurium. Como en el caso de gdhA, hay una homología importante entre el ARNt de histidina y la secuencia de la estructura tipo ARNt. Ellos han propuesto que esta estructura tipo ARNt pudiera tener alguna función regulatoria. Aun más, secuencias tipo ARNt han sido encontradas en algunos virus eucariotes de ARN y estos son reconocidos y aminoacilados por las enzimas del huésped, a pesar de una desviación significativa de la estructura consenso del ARNt (Haenni et al., 1982).

Un papel alternativo interesante para las secuencias REP como sitios específicos de unión de proteínas, se deriva del hecho de que las secuencias tipo REP que han sido encontradas en los ARNs ribosomales 16 y 23S, ya han sido involucradas en la unión de proteínas ribosomales específicas. La secuencia tipo REP del ARN ribosomal 16S, interacciona con la proteína S7 (Prince et al., 1983;

Noller, 1984); y en el caso del ARN ribosomal 23S, la secuencia tipo REP es parte de un dominio estructural que interacciona con las proteínas L7, L10, L11, L12 y el factor de elongación G (Prince et al., 1983). Por lo tanto, un posible papel interesante para las secuencias REP ((no incluida entre las posibles funciones alternativas propuestas por Stern et al., (1984)), pudiera ser que estas secuencias nucleotídicas interaccionaran con proteínas específicas, no necesariamente nucleasas sino proteínas de la maquinaria general de traducción de la célula. Esto podría explicar porqué ellas se encuentran no solamente en moléculas de ARN mensajero, sino también en ARNs estables tales como los ARNs ribosomales 16 y 23S o en moléculas de ARNt. Nos gustaría sugerir que mediante la interacción de estas secuencias REP con enzimas modificadoras específicas y/o proteínas ribosomales específicas, se pudiera modular la velocidad de la traducción y/o degradación de ciertas moléculas de ARN mensajero.

Otro papel alternativo que también ha sido propuesto por Stern et al., (1984), es la posible participación de las secuencias REP en la estructuración y organización del cromosoma, posiblemente sirviendo como sitios de unión para proteínas específicas. Tales proteínas pudieran anclarse al ADN o pudieran servir para delinear dominios estructurales o funcionales. Sin embargo, esta propuesta puede estructurarse de una manera diferente: ha sido demostrado que las secuencias REP forman parte de aquellos transcritos en los que se las ha buscado, sugiriendo que la transcripción de estas secuencias es importante (Becerril et al., 1985).

Worcel y Burgi, (1972), han demostrado que las fuerzas que mantienen la estructuración y estabilizan la arquitectura del

cromosoma bacteriano enrollado, se deben fundamentalmente a las interacciones entre ADN y ARN. Por tanto, es posible que la transcripción de las secuencias REP pudiera contribuir a estas interacciones.

Finalmente, también se ha sugerido que estas secuencias están relacionadas con la evolución del genoma bacteriano (Gilson *et al.*, 1984; Stern *et al.*, 1984). Gilson *et al.*, han observado cierta homología entre las secuencias REP y la secuencia del extremo 5' de las secuencias de inserción IS10R e IS10L así como con algunos transposones. Se argumenta que las secuencias REP podrían promover rearrreglos entre regiones distantes del genoma. Adicionalmente, Stern *et al.*, han obtenido algunas evidencias acerca de la recombinación entre las secuencias REP presentes en plásmido y cromosoma. Estos autores sugieren que estas secuencias pudieran promover rearrreglos genómicos que afectarían tanto la expresión como la evolución de los genes.

PERSPECTIVAS.

A continuación se puntualizan algunos aspectos que serán tomados en consideración en las investigaciones futuras sobre el gene gdhA:

En lo que respecta a la región flanqueadora 5', se planea hacer uso de mutantes deficientes en la adenilato ciclasa (cya⁻) y/o CRP (crp⁻), con el objeto de demostrar la influencia de la proteína CRP sobre la expresión del gene.

Por otro lado, se harán mapeos de promotor en condiciones metabólicas diferentes a las que ya han sido probadas, con la intención de conocer si sólo funciona el promotor ya mapeado o si existen otros promotores funcionales.

También, se harán estudios de la posible influencia de las secuencias aledañas al sitio de TaqI, sobre el doblamiento del ADN, tema que cobra cada día una mayor atención en los estudios de la importancia de la arquitectura del ADN en la posible regulación de los genes.

En lo referente a la región flanqueadora 3' se contempla precisar el efecto de las deleciones de las secuencias REP sobre la estabilidad del mensajero de la GDH.

Otro punto relacionado que se estudiará es la posible participación de la ARNasa III en el procesamiento del mensajero de la GDH.

Por otra parte, se tiene previsto estudiar la posible participación de las secuencias REP en la estructuración del cromosoma.

Se planea también estudiar dentro de lo posible, que tan ampliamente distribuidas se encuentran las secuencias REP entre las bacterias, y especular sobre sus posibles implicaciones evolutivas.

REFERENCIAS

- Adhya, S. and Gottesman, M.: Control of transcription termination. *Annu. Rev Biochem.* 47 (1978) 967-996.
- Ames, B.N., Tsang, T.H., Buck, M. and Christman, M.F.: The leader of the Histidine attenuator region resembles tRNA His. Possible general regulatory implications. *Proc. Natl. Acad. Sci. USA* 80 (1983) 5240-5242.
- Ashby, B., Nixon, J.S. and Wooton, J.C.: Mutational variants of the Neurospora crassa NADP-specific Glutamate Dehydrogenase altered in a conformational equilibrium. *J. Mol. Biol.* 149 (1981) 521-540.
- Austen, B.M., Haberland, M.E. and Smith, E.L.: Secondary structure predictions for the NAD-specific Glutamate Dehydrogenase of Neurospora crassa. *J. Biol. Chem.* 255 (1980) 8001-8004.
- Ausubel, F.M.: Regulation of Nitrogen Fixation Genes. *Cell* 37 (1984) 5-6.
- 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.
- Blumenthal, K.M., Moon, K. and Smith, E.L.: Nicotinamide adenine dinucleotide Phosphate-specific Glutamate Dehydrogenase of Neurospora. *J. Biol. Chem.* 250 (1975) 3644-3654.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Falkow, S.: Construction and characterization of new cloning vehicles, II. A multipurpose cloning system. *Gene* 2 (1977) 95-113.
- Brenchley, J.E., Prival, M.J. and Magasanik, B.: Regulation of the synthesis of enzymes responsible for glutamate formation in Klebsiella aerogenes. *J.*

- Biol. Chem. 248 (1973) 6122-6128.
- Brett, M. Chambers, G.K., Holder, A.A., Fincham, J.R.S. and Wootton, J.C.:
Mutational aminoacid replacement in Neurospora crassa NADP-specific
Glutamate Dehydrogenase. J. Mol. Biol. 106 (1976) 1-22.
- Cid, H., Bunster, M., Arriagada, E. and Campos, M.: Predictions of secondary
structure of proteins by means of hydrophobicity profiles. FEBS Lett. 150
(1982) 247-254.
- Clarke, P.H. Ed: Biological Regulation and Development. (Goldberger, Ed).
Plenum Press. New York (1979) 145-149.
- Coulton, J.W. and Kapoor, M.: Studies on the kinetics of Glutamate
Dehydrogenase of Salmonella typhimurium. Can. J. Microbiol. 19 (1973)
439-450.
- Covarrubias, R.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.
- de Bruijn, F.J. and Ausubel, F.M.: The cloning and characterization of the
glnF (ntrA) gene of Klebsiella pneumoniae: Role of glnF(ntrA) in the
regulation of nitrogen fixation (nif) and other nitrogen fixation genes.
Mol. Gen. Genet. 192 (1983) 342-353.
- Galas, D.J., Eggert, M. and Waterman, S.M.: Analysis of promoters of
Escherichia coli. Rigorous pattern-recognition methods for DNA sequences.
J. Mol. Biol. 186 (1985) 117-128.
- 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.

- Gouy, M. and Gautier, C.: Codon usage in bacteria: Correlation with gene expressivity. *Nucl. Acids Res.* 10 (1982) 7055-7074.
- Guldi-Rontani, C., Danchin, A. and Ullman, A.: Transcriptional control of polarity in Escherichia coli by cAMP. *Mol. Gen. Genet.* 195 (1984) 96-100.
- 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.
- Grosjean, H. and Fiers, W.: Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 18 (1982) 199-209.
- Haenni, A., Joshi, S. and Chapeville, F. En Cohn, W.E. (Ed.), *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, New York (1982) 85-102.
- Haberland, M.E., Chen, C.W. and Smith, E.L.: NAD-specific Glutamate Dehydrogenase of Neurospora crassa. *J. Biol. Chem.* 255 (1980) 7993-8000.
- Hawley, D.K. and McClure, W.R.: Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucl. Acids Res.* 11 (1983) 2237-2255.
- Heidecker, G., Messing, J. and Gronenborn, B.: A Versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* 10 (1980) 69-73.
- Higgins, C.F., Ames, F.L.G., Barnes, W.M., Clement, J.M. and Hofnung, J.M.: A novel intercistronic regulatory element of prokaryotic operons. *Nature* 298 (1982) 760-762.
- Hirschman, J., Wong, P.K., Sei, K., Keener, J. and Kustu, S.: Products of nitrogen regulatory genes ntxA and ntxC of enteric bacteria activate glnA transcription in vitro: Evidence that the ntxA product is a σ factor. *Proc. Natl. Acad. Sci. USA* 82 (1985) 7525-7529.
- Ho, Y.S., Wulff, D.L. and Rosenberg, M.: Bacteriophage λ protein cII binds promoters on the opposite face on the DNA helix from RNA polymerase. *Nature* 304 (1985) 703-708.

- Holmes, W.M., Platt, T. and Rosenberg, M.: Termination of transcription in E. coli. Cell 32 (1983) 1029-1032.
- Hunt, T.P. and Magasanik, B.: Transcription of glnA by purified Escherichia coli components: Core RNA Polymerase and the products of glnF, glnG and glnL. Proc. Natl. Acad. Sci. USA 82 (1985) 8453-8457.
- Julliard, J.H. and Smith, E.L.: Partial amino acid sequence of the Glutamate Dehydrogenase of human liver and a revision of the sequence of the bovine enzyme. J. Biol. Chem. 254 (1974) 3427-3438.
- Kustu, S., Burton, D., Garcia E., McCarter, L. and McFarland, N.: Nitrogen control in Salmonella: Regulation by glnR and glnF gene products. Proc. Natl. Acad. Sci. USA 76 (1979) 4576-4580.
- León, P., Romero, D., Garcíafarrubio, A., Bastarrachea, F. and Covarrubias, A. A.: Glutamine synthetase-constitutive mutation affecting the glnALB upstream promoter of Escherichia coli. J. Bact. 164 (1985) 1032-1038.
- MacNeil, T., Roberts, G.P., MacNeil, D. and Tyler, B.: The products of glnL and glnG are bifunctional regulatory proteins. Mol. Gen. Genet. 188 (1982) 325-333.
- Magasanik, B. and Stadtman, E.: Regulation of glutamine synthetase. En: Biological Recognition and Assembly. Alan R. Liss, Inc., New York (1980) 337-343.
- Marini, J.C., Levene, S.D., Crothers, D.M. and Englund, P.T.: A bent helix in kinetoplast DNA. Cold Spring Harbor Symposia on Quantitative Biology. Vol. XLVII parte 1 (1983) 279-283.
- McClure, W.R.: Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. 54 (1985) 171-204.
- McFarland, N., McCarter, L., Artz, S. and Kustu, S.: Nitrogen regulatory locus "glnR" of enteric bacteria is composed of cistrons ntxB and ntxC: Identification of the protein products. Proc. Natl. Acad. Sci. USA 78

(1981) 2135-2139.

- Merrick, M.J.: A new model for nitrogen control. *Nature* 297 (1982) 362-363.
- Messing, J., Crea, R. and Seeburg, P.H.: A system for shotgun DNA sequencing *Nucl. Acids Res.* 9 (1981) 309-321.
- Musso, R., Di Lauro, S., Adhya, S. and de Crombrughe, P.: Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. *Cell* 12 (1977) 847-854.
- Noller, H.F.: Structure of ribosomal RNA. *Annu. Rev. Biochem.* 53 (1984) 119-162.
- Dw, D.W., Sundaresan, V., Rothstein, D.M., Brown, S.E. and Ausubel, F.M.: Promoters regulated by the glnG (ntrC) and nifA gene products share a heptameric consensus sequence in the -15 region. *Proc. Natl. Acad. Sci. USA* 80 (1983) 2524-2528.
- Pahel, G., Zelenetz, A. and Tyler, B.: gltB gene and regulation of nitrogen metabolism by glutamine synthetase in Escherichia coli. *J. Bact.* 133 (1978) 134-148.
- Pahel, G. and Tyler, B.: A new glnA linked regulatory gene for glutamine synthetase in Escherichia coli. *Proc. Natl. Acad. Sci. USA* 76 (1979) 4544-4548.
- Plaman, M.D. and Stauffer, C.V.: Characterization of a cis-acting regulatory mutation that maps at the distal end of the Escherichia coli glyA gene. *J. Bact.* 161 (1985) 650-654.
- 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.
- Prival, M.J., Brenchley, J.E. and Magasanik, B.: Glutamine synthetase and the regulation of histidase formation in Klebsiella aerogenes. *J. Biol. Chem.* 248 (1973) 4334-4344.
- Prussiner, S., Miller, R.E. and Valentine, R.C.: Adenosine 3'5' cyclic

- monophosphate control of the enzymes of glutamine metabolism in Escherichia coli. Proc. Natl. Acad. Sci. USA 69 (1972) 2922-2926.
- Rosenberg, M. and Court, D.: Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13 (1979) 319-353.
- Sakamoto, N., Kotre, A.M. and Savageau M.A.: Glutamate dehydrogenase from Escherichia coli: Purification and properties. J. Bact. 124 (1975) 775-783
- Sánchez-Pescador, R., Sanvicente, E., Valle, F. and Bolivar, F.: Recombinant plasmids carrying the glutamate dehydrogenase gene from Escherichia coli K-12. Gene 17 (1982) 1-8.
- Sanvicente, E., Sanchez-Pescador, R., Valle, F. y Bolivar, F.: Evidencias Bioquímicas de la presencia del gene estructural de la deshidrogenasa glutámica de E. coli K-12. Bol. Est. Med. Biol. 32 (1983) 225-232.
- Senior, P.J.: Regulation of nitrogen metabolism in Escherichia coli and Klebsiella aerogenes: Studies with the continuous culture technique. J. Bact. 123 (1975) 407-418.
- Smith, F., Austen, B., Blumenthal, K.M. and Nyc, J.: Glutamate Dehydrogenase En: The Enzymes (Boyer, P.O., ed). Vol. XI Academic Press, New York (1975) 293-367.
- 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 mayor component of the bacterial genome. Cell 37 (1984) 1015-1026.
- Stewart, G.R., Mann, A.F. and Fenten P.A.: Enzymes of glutamate formation: Glutamate dehydrogenase, glutamine synthetase and glutamate synthase. En: The Biochemistry of Plants. (Hiplin, E.J. ed) Academic Press, New York, Vol 5 (Aminoacids and Derivatives) (1980).
- Streitcher, S.L., De Leo, A.B. and Magasanik, B.: Regulation in Klebsiella

- aerogenes by episomal glutamine synthetase of Escherichia coli. J. Bact. 127 (1976) 184-192.
- 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.
- Valle, F.: Estudio de las regiones de regulación del gene estructural de la enzima glutamato deshidrogenasa de Escherichia coli K-12. Tesis Maestría UACPyP-CCH, UNAM, México, D.F. (1984)
- Varrichio, F.: Control of glutamate dehydrogenase synthesis in Escherichia coli. Biochem. Biophys. Acta. 177 (1969) 560-564.
- Veronesse, F.M., Boccu, E. and Conventi, L.: Glutamate dehydrogenase from Escherichia coli: Induction purification and properties of the enzyme. Biochem. et Biophys. Acta. 377 (1975) 217-228.
- Wald, G.: Life in the second and third periods; or why phosphorous and sulfur for high energy bonds?. En: Horizons in Biochemistry. (Kasha, M. and Pullman, B., eds.) (1962) 127-142.
- Windass, J.D., Worsey, M.J., Pioli, E.M., Pioli, D., Barth, P.T., Atherton, K.T., Dart, E.C., Byron, D., Powell, K. and Senior, P.J.: Improved conversion of methanol to single cell protein by Methylophilus methylotropus. Nature 287 (1980) 346-401.
- Worcel, A. and Burgi, E.: On the structure of the folded chromosome of Escherichia coli. J. Mol. Biol. 71 (1972) 127-147.

Wu, H.M. and Crothers, D.M.: The locus of sequence-directed and protein induced DNA bending. *Nature* 308 (1984) 509-513.

Yanofsky, C.: Attenuation in the control of expression of bacterial operons. *Nature* 289 (1981) 751-758.

GENE 951

Complete nucleotide sequence of the glutamate dehydrogenase gene from *Escherichia coli* K-12[†]

(Nitrogen metabolism; protein homology; codon usage; DNA sequence)

Fernando Valle, Baltazar Becerril, Ellison Chen*, Peter Seeburg*, Herbert Heyneker*, and Francisco Bolivar**

Departamento de Genética y Biología Molecular, Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, México 04510, D.F. (México) Tel. (915) 550-3893, and * Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080 (U.S.A.) Tel. (415) 952-1000

(Received October 15th, 1983)

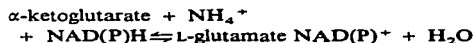
(Accepted November 3rd, 1983)

SUMMARY

A 2.3-kb *Pst*I-*Cla*I chromosomal DNA segment, carrying the complete coding region of the glutamate dehydrogenase (GDH) structural gene from *Escherichia coli* K-12, has been sequenced. The complete amino acid sequence (447 residues) of the GDH monomer has been deduced, and comparisons are made with reported amino acid sequences of GDH from other organisms.

INTRODUCTION

L-Glutamate dehydrogenases (EC 1.4.1.2-4), as a class, catalyze the interconversion of α -ketoglutarate and L-glutamic acid:



These enzymes provide a route for incorporation of nitrogen into organic compounds, and thus a link between carbohydrate and amino acid metabolism.

There are at least three types of GDHs, which differ in coenzyme specificity: those specific for

either NADH or NADPH and those that can function with both. It is essential to note, however, that these enzymes vary not only in coenzyme specificity but also in other properties, e.g., induction and repression of synthesis by metabolites, regulation of activity by purine nucleotide di- and triphosphate (ADP, GDP, ATP, GTP), and other ligands and also in molecular properties (Smith et al., 1975). A number of reports have appeared in the literature which accentuated these differences, and it is now clear that the role these enzymes play in metabolic processes may vary from organism to organism.

GDHs show an unusually strong degree of sequence conservation, as demonstrated by amino acid sequence comparisons of various purified enzymes from different organisms (Julliard and Smith, 1979). Also, strong similarities have been reported between the prokaryotic *E. coli* and the eukaryotic *Neurospora crassa* NADP-dependent GDHs (Matta et al., 1982; Valle et al., 1983).

[†] Dedicated to the memory of Ahmad I. Bukhari.

** To whom correspondence and reprint requests should be mailed.

Abbreviations: bp, base pairs; GDH, glutamate dehydrogenase; kb, kilobases or kilobase pairs.

We have previously reported the nucleotide sequence of the regulatory and amino-terminal coding regions of the *E. coli* *gdhA* gene (Valle et al., 1983). In this paper, we report the complete nucleotide sequence of this gene and its flanking regions. Amino acid sequence comparisons between the *E. coli* GDH deduced sequence and the enzyme sequences from other organisms are also presented.

MATERIALS AND METHODS

(a) Bacterial strains, plasmids and phages

Bacteriophage M13mp7 and its permissive host JM101 [*thi, supE Δ(proAB-lac) F' (traD36, proAB, lacZM15, lacI^h)*] were obtained from J. Messing. Plasmid pSAE4 has been previously described (Sánchez-Pescador et al., 1982; Valle et al., 1983).

(b) Enzymes

Restriction enzymes *TaqI*, *HpaI*, *HincII* and BRL, Inc. *PstI*, *BamHI* purified according to Greene et al. (1978) and

(c) Nucleotide sequence

Procedures described by Messing et al. (1980) for nucleotide sequences of DNA fragments, derived into the M13mp7 with *HincII*, *AclI*, *AclI*

RESULTS

(a) Nucleotide sequence K-12

A detailed restriction fragment from plasmid *E. coli* K-12 is presented in Fig. 2. The DNA fragments from the phage M13mp7 as (1980). A collection clones was used to presented in Fig. 2. The sequence of the non-coding and its flanking region the GDH monomer sequence of the gene deduced M_r of the enzyme 48 589.5 and is in agreement with values based on reported (Sakamoto et al., 1981)

(b) GDH amino acid

A comparison of the deduced sequence with the corresponding amino acid sequences of Julliard and Smith

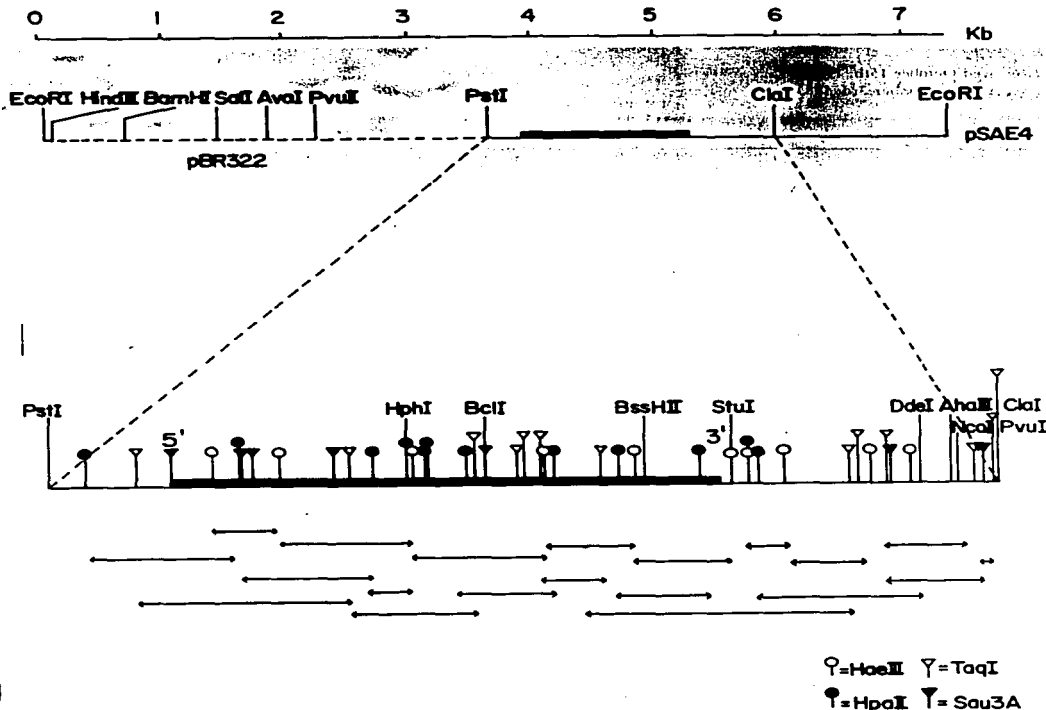


Fig. 1. Physical map of the linear form (*EcoRI* cut) of plasmid pSAE4. The horizontal dashed line represents the vector pBR322, and the thick bar represents the location of the *gdhA* gene with 3' and 5' ends indicated at the lower part. Different DNA fragments from the plasmid were subjected to sequence analysis with the strategy indicated in the lower portion of the figure.

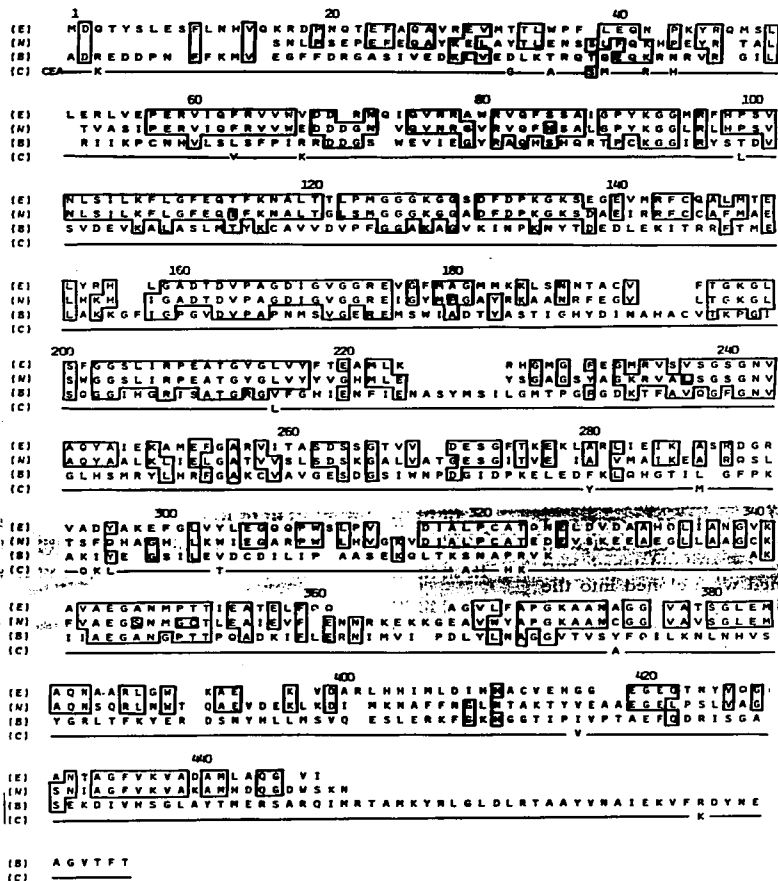


Fig. 3. Comparison of amino acid sequences of the GDH monomer from *E. coli* (E), *N. crassa* (N), bovine (B) and chicken liver (C). Numbering refers to the *E. coli* sequence. Data for sources other than *E. coli* are from Blumenthal et al. (1975). Identical amino acid sequences are boxed. The sequence of the chicken enzyme (solid line) is identical to that of the bovine, except at the positions where specific substitutions are shown.

Table I

Codon usage frequency in the

Arg	CGC - 12
	CGT - 9
	CGA - 1
	CGG - 0
	AGG - 0
	AGA - 0
Leu	CTG - 18
	CTT - 6
	CTA - 3
	TTA - 1
	CTC - 6
	TTG - 1
Ser	AGC - 6
	TCA - 5
	TCT - 4
	TCC - 5
	AGT - 1
	TCG - 1

mately the same number of two NADP-dependent GDH monomers as other than to the monomers.

(c) Codon usage frequency

Table I shows the codon usage of the *gdhA* gene. As can be seen, the codon usage is similar to that of other genes which are efficiently translated (Gouy and Gautier,

DISCUSSION

We have reported the amino acid sequence of the regulon region of the *gdhA* gene from *N. crassa* (Blumenthal et al., 1983). We have shown that the 5' end of the gene contains a putative ribosome binding site. Published results indicate that under high

Table I

Codon usage frequency in the *gdhA* gene from *E. coli* K-12

Arg	CGC - 12 CGT - 9 CGA - 1 CGG - 0 AGG - 0 AGA - 0	Ala	GCC - 16 GCA - 10 GTC - 8 GCG - 13	Val	GTG - 9 GTC - 8 GTT - 16 GTA - 4	Glu	GAA - 23 GAG - 8
Leu	CTG - 18 CIT - 6 CTA - 3 TTA - 1 CTC - 6 TTG - 1	Gly	GGC - 22 GGT - 21 GGG - 4 GGA - 2	Ile	ATC - 12 ATT - 4 ATA - 1	His	CAT - 6 CAC - 4
Ser	AGC - 6 TCA - 5 TCT - 4 TCC - 5 AGT - 1 TCG - 1	Pro	CCG - 12 CCC - 0 CCT - 2 CCA - 1	Asn	AAC - 9 AAT - 8	Lys	AAA - 18 AAG - 4
		Thr	ACC - 11 ACT - 6 ACA - 4 ACG - 1	Asp	GAT - 14 GAC - 5	Phe	TTC - 11 TTT - 11
				Cys	TGC - 3 TGT - 1	Tyr	TAT - 5 TAC - 5
				Gln	CAG - 15 CAA - 8	Met	ATG - 18
						Trp	TGT - 5

mately the same number of residues (450), while the two NADP-dependent monomers are composed of 501 residues. Furthermore, the *E. coli* and *N. crassa* GDH monomers are more homologous to each other than to the NADH-dependent GDH monomers.

(c) Codon usage frequency

Table I shows the codon usage frequency of the *gdhA* gene. As can be seen, there is preferential codon usage, similar to those genes whose mRNAs are efficiently translated (Grosjean and Fiers, 1982; Gouy and Gautier, 1982).

DISCUSSION

We have reported the cloning and nucleotide sequence of the regulatory and amino-terminal coding region of the *gdhA* gene from *E. coli* K-12 (Valle et al., 1983). We have described the presence, at the 5' end of the gene, of three putative promoters and a putative ribosome-binding site for this gene. Unpublished results, using S1 nuclease mapping, indicate that under high-ammonium growth conditions,

transcription is initiated from putative promoter P3 (Valle, F., Becerril, B. and Bolivar, F., manuscript in preparation). Another interesting characteristic of this gene is that it lacks the typical transcriptional attenuation regulatory features present in various amino acid biosynthetic operons (Yanofsky, 1981).

The nucleotide sequence of the *gdhA* gene, presented in Fig. 2, shows an open reading frame containing 1341 bp starting with an ATG at position +1 and ending with a TAA triplet. This region can encode for a polypeptide chain 447 amino acids long. Examination of the DNA sequence at the 3' end of the structural gene (approx. 200 bp) reveals an RNA transcript carrying this region with the capacity to form different secondary structures. One of them, a stable stem and loop structure 87 nucleotides after the UAA stop codon, has the general characteristics of a rho-independent terminator (Adhya and Gottesman, 1978; Holmes et al., 1983).

Higgins et al. (1982) have recently proposed a novel regulatory element of prokaryotic genes. This element consists of a long dyad symmetry which, once transcribed, could form an exceptionally stable stem-loop structure with a stem of up to 36 bp. In this region, smaller palindromic units some of which overlap with the main symmetry, are also found. The nucleotide sequence of this element is greatly con-

served and has been localized at intercistronic regions of several *E. coli* operons and also following the *trpR* gene. In this regard it is interesting to mention that, beginning at position 1352 (8 bp after the TAA codon) of the *gdhA* gene, it is possible to find the same kind of structure with remarkable homology (28 of 29 identities) with the consensus sequence proposed by Higgins et al. (1983) (Fig. 2). Experiments designed to determine whether this region acts in the regulation of this gene are currently underway. It is also important to mention that analysis of approx. 800 bp at the 3' region of the *gdhA* structural gene did not reveal any long open reading frame in any of the three phases. This suggests that the *gdhA* gene of *E. coli* is not part of a classical bacterial operon.

The amino acid sequence of the *E. coli* glutamate dehydrogenase (NADP-dependent) monomer reported here is compared, in Fig. 3, with the sequences of the GDH monomers from *N. crassa*, and bovine and chicken liver. To obtain maximal homology, it was necessary to introduce small gaps in the sequences. Within the limits imposed by these alignments, we can see excellent sequence conservation among the GDH monomers from *E. coli* and *N. crassa*. Less homology is observed between the bovine and *E. coli* proteins. Examination of these data indicates that many of these residues are probably conserved for conformational importance as opposed to catalytic effects (Blumenthal et al., 1975). For example, as pointed out by Blumenthal et al. (1975), the majority of the conserved glycine residues between these GDH monomers must be important for conformation, since it is well known that glycine residues are generally present in nonhelical portions of the protein structure.

The greatest homology among the four proteins is found in the amino-terminal half of the monomers. This section of the protein, which includes a highly reactive lysine residue (Blumenthal et al., 1975) at the amino acid position 128 of the *E. coli* sequence, probably includes the residues involved in the binding of α -ketoglutarate, and in catalytic activity. These propositions are supported by chemical modification studies and the properties of mutational variants (Brett et al., 1976; Julliard and Smith, 1979; Haberland and Smith, 1980; Austen et al., 1980; Ashby et al., 1981).

In addition to the observed identities, a number of

substitutions in the sequence conserve the same character of bulk hydrophobicity and accessibility coefficient (Cid et al., 1982). These observations suggest that these proteins could be even more homologous to each other at the tridimensional level than in their amino acid sequence. In this regard, using the method described by Cid et al. (1982) for the prediction of protein secondary structure, we have found that in some regions, where the amino acid sequence is not conserved between the *E. coli* and the bovine GDH monomers (but conserved between *E. coli* and *N. crassa*), both proteins could form the same secondary structure (data not shown).

Table I shows the codon usage frequency of the *E. coli* *gdhA* gene. As can be seen, there are certain codons that are infrequently utilized and others such as CGG, CCC, AGA and AGG that are not used at all. It has been proposed that some of these codons may act as modulators of mRNA translation in *E. coli* (Grosjean and Fiers, 1982). The absence of these codons in the *gdhA* gene of *E. coli* may indicate the existence of strong evolutionary pressure for maximal expression of this gene. Ikemura (1981) has proposed a balance between selection for optimal codons and the mutational process such that, when a gene does not have a high translational rate, selective pressure is weak and the mutational process may blur its effect. On the other hand, for highly expressed genes, the pressure is stronger and optimization occurs. However, the notion that the *gdhA* gene is under a strong evolutionary pressure is not supported by the fact that this gene does not appear to be absolutely necessary for the growth of *E. coli* cells. It has been shown that glutamate can be synthesized by the glutamate synthase enzyme, and *gdh⁻* cells do not appear to be at a metabolic disadvantage.

REFERENCES

- Adhya, S. and Gottesman, M.: Control of transcription termination. *Annu. Rev. Biochem.* 47 (1978) 967-996.
- Ashby, B., Nixon, J.S. and Wootton, J.C.: Mutational variants of all the *Neurospora crassa* NADP-specific glutamate dehydrogenase altered in a conformational equilibrium. *J. Mol. Biol.* 149 (1981) 521-540.
- Austen, B.M., Haberland, M.E. and Smith, E.L.: Secondary structure predictions for the NAD-specific glutamate dehydrogenase of *Neurospora crassa*. *J. Biol. Chem.* 255 (1980) 8001-8004.

- Blumenthal, K.M., Moon, adenine dinucleotide plus
genase of *Neurospora*. J.
Brett, M., Chambers, G.K.,
Wootton, J.C.: Mutant
Neurospora crassa NAD
J. Mol. Biol. 106 (1976)
Cid, H., Bunster, M., Arraiz
of secondary structure
city profiles. *FEBS Lett*
Gouy, M. and Gautier, C.:
with gene expressivity. J
Greene, P.J., Heyncker,
Betlach, M.C., Covart
D.J., Tait, R. and Roy
purification of restrict
2373-2380.
Grosjean, H. and Fiers, W.
ryotic genes: the opti
the selective codon
Gene 18 (1982) 199-2
Haberland, M.E., Chen
glutamate dehydrog
Chem. 255 (1980) 799
Heidecker, G., Messing,
primer for DNA sequ
Gene 10 (1980) 69-75
Higgins, C.T., Ames, G.
Hofaung, M.: A new
prokaryotic operons.
Holmes, W.M., Platt, T
transcription in *E. co*
Ikemura, T.: Correlation
coli transfer RNAs
codons in its prot

nc—conserve the same
ob ty and accessibility
(2). These observations
could be even more
the tridimensional level
quence. In this regard,
by Lid et al. (1982) for
ec dary structure, we
editions, where the amino
ved between the *E. coli*
tors (but conserved
), both proteins could
cture (data not shown).
use e frequency of the
sion, there are certain
utilized and others such
GC that are not used at
sion of these codons
mRNA translation in
1987). The absence of
of *E. coli* may indicate
ationary pressure for
e. Temura (1981) has
selection for optimal
cess such that, when
nstructional rate, selec-
tional process may
id, for highly expressed
gen and optimization
the *gdhA* gene is
sure is not supported
as not appear to be
wt of *E. coli* cells. It
can be synthesized by
and *gdh⁻* cells do
d advantage.

- Blumenthal, K.M., Moon, K. and Smith, E.L.: Nicotinamide adenine dinucleotide phosphate-specific glutamate dehydrogenase of *Neurospora*. *J. Biol. Chem.* 250 (1975) 3644-3654.
- Brett, M., Chambers, G.K., Holder, A.A., Fincham, J.R.S. and Wootton, J.C.: Mutational amino acid replacements in *Neurospora crassa* NADP-specific glutamate dehydrogenase. *J. Mol. Biol.* 106 (1976) 1-22.
- Cid, H., Bunster, M., Arriagada, E. and Campos, M.: Prediction of secondary structure of proteins by means of hydrophobicity profiles. *FEBS Lett.* 150 (1982) 247-254.
- Gouy, M. and Gautier, C.: Codon usage in bacteria: correlation with gene expressivity. *Nucl. Acids Res.* 10 (1982) 7055-7074.
- Greene, P.J., Heyneker, H.L., Bolivar, F., Rodriguez, R.L., Betlach, M.C., Covarrubias, A.A., Backman, K., Russel, D.J., Tait, R. and Boyer, W.H.: A general method for the purification of restriction enzymes. *Nucl. Acids Res.* 5 (1978) 2373-2380.
- Grosjean, H. and Fiers, W.: Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction and the selective codon usage in efficiently expressed genes. *Gene* 18 (1982) 199-209.
- Haberland, M.E., Chen, C.W. and Smith, E.L.: NAD-specific glutamate dehydrogenase of *Neurospora crassa*. *J. Biol. Chem.* 255 (1980) 7993-8000.
- Heidecker, G., Messing, J. and Gronenborn, B.: A versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* 10 (1980) 69-73.
- Higgins, C.T., Ames, G.F., Barnes, W.M., Clement, J.M. and Hofaung, M.: A novel intergenic regulatory element of prokaryotic operons. *Nature* 298 (1982) 760-762.
- Holmes, W.M., Platt, T. and Rosenberg, M.: Termination of transcription in *E. coli*. *Cell* 32 (1983) 1029-1032.
- Ikemura, T.: Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151 (1981) 389-409.
- Julliard, J.H. and Smith, E.L.: Partial amino acid sequence of the glutamate dehydrogenase of human liver and a revision of the sequence of the bovine enzyme. *J. Biol. Chem.* 254 (1979) 3427-3438.
- Mattaaj, I.W., McPherson, M.J. and Wootton, J.C.: Localization of a strongly conserved section of coding sequence in glutamate dehydrogenase genes. *FEBS Lett.* 147 (1982) 21-25.
- Messing, J., Crea, R. and Seeburg, P.H.: A system for shotgun DNA sequencing. *Nucl. Acids Res.* 9 (1981) 309-321.
- Sakamoto, N., Kotre, A.M. and Savageau, M.A.: Glutamate dehydrogenase from *Escherichia coli*: Purification and properties. *J. Bacteriol.* 124 (1975) 775-783.
- Sánchez-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.
- Smith, F., Austen, B., Blumenthal, K.M. and Nyc, J.: Glutamate dehydrogenase, in Boyer, P.D. (Ed.) *The Enzymes* Vol. XI. Academic Press, New York, 1975, pp. 293-367.
- Tait, R.C., Rodriguez, R.L. and West, R.W.: The rapid purification of T4 DNA ligase from λ T4lig lysogen. *J. Biol. Chem.* 255 (1980) 813-814.
- 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.
- Yanofsky, C.: Attenuation in the control of expression of bacterial operons. *Nature* 289 (1981) 751-758.

Communicated by R.L. Rodriguez.

of transcription termi-
78) 87-996.
Mutational variants of
glutamate dehydrogen-
1981). *J. Mol. Biol.* 149

Smith, E.L.: Secondary
specific glutamate dehy-
Biol. Chem. 255 (1980)

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 Riba and Francisco Bolivar*

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^- , *lacS*20 ($r_B^- m_B^-$), *recA*13, *ara*-14, *proA*2, *lacY*1, *galK* $^-$, *rpsL*20 (Sm^R), *xyl*-5, *mit*-1, *supE*44, λ^-] and the glutamate auxotroph PA340 (F^- , *leuB*6, *gdh*-1, *hisG*1, *gltB*31, *argH*1, *thi*-1, *ara*-14, *lacY*1, *galE*, *malA*1, *xyl*-7, *met*-2, *tonA*2, λ^R , λ^- , *rpsL*9, *supE*4) were used to transform plasmid constructions as described by Cohen et al. (1972). Plasmid pSAE⁺ has been previously described (Sanchez-Pescador et al., 1982; Valle et al., 1983).

(b) Enzymes

*Stu*I and *Nru*I 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 *Stu*I or *Stu*I + *Nru*I, 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

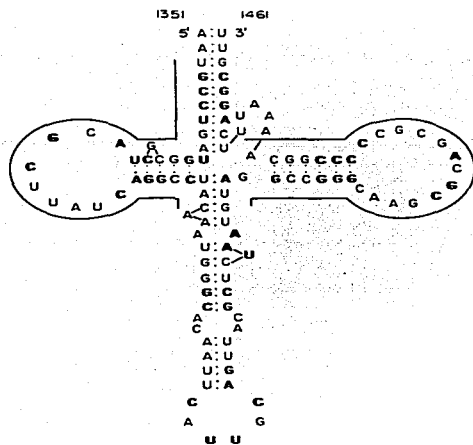


Fig. 2. A tRNA-like structure resulting from folding of the RNA transcript carrying the first 120 nt after TAA stop codon (for numbering see Fig. 1). The nt in common with glutamate tRNA appear in bold characters. The location of REP sequences are indicated by solid lines.

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 *SruI* sites located at the 3' region outside the coding region of the *gdhA* gene (positions 1375 and 2375, Fig. 1), and a unique *NruI* site located in the Tc^R gene of the cloning vehicle pBR322 (Belivar 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*,

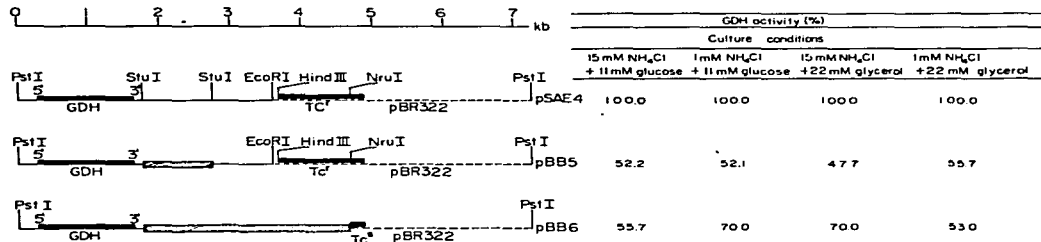


Fig. 3. Physical maps of the plasmids pSAE4, pBB5 and pBB6 linearized at the *PstI* site. The dashed lines represent the vector pBR322, the solid lines represent *E. coli* chromosomal inserts and the thick bars indicate the locations of the *gdhA* and Tc^R genes. The hatched areas indicate deleted regions. 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 Higgings 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*, *malE*, *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>aroG</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	—
M1 RNA		c	3,5
<i>spot42</i>		60	25
<i>hisJ</i> (S.t)		47	20
<i>hisG</i> (S.t)		6	2,5
<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>araBAD</i> (S.t)		14 (f)	6,11

^a *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 intercistronic 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).

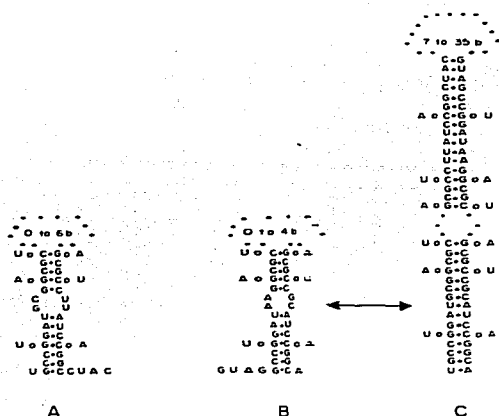


Fig. 4. Stem and loop structures that can be potentially formed within the REP consensus sequences. Structures A and B, were generated from consensus sequences presented in Tables II and III, respectively. Structure C was generated by pairing structures A and B. Note that the paired regions in A and B are identical; conserved sequences CUAC and GUA, found at the 3' end in A and at the 5' end in B, and also the central bubble on each structure, confer upon them directionality with respect to the proximal gene. It is important to point out that the REP consensus sequence proposed by Stern et al. (1984), is the same as the REP consensus sequence in this figure (A and B), except that there is an additional uridine at the 5' end of A and an additional adenine at the 3' end of B.

and of growth medium (Fig. 3). This effect seems to be independent of the sequences located downstream of the deleted REP sequence (see pBB5 vs. pBB6; Fig. 3).

The detailed analysis of the GenBank database has allowed us to increase the number of reported REP sequences from 35 to 51 and the number of transcriptional units where these sequences are present from 16 to 25. More information was obtained from this analysis. The consensus REP sequence was increased from 32 to 33 bp (Tables II and III), and we also found that, in the consensus REP sequences, the loops have a restricted size, 0-4 or 0-6 bp. Another interesting piece of information, is that in 2 of the 9 new transcriptional units reported in this work, in which REP has been located, these types of sequence have been found overlapping regulatory regions, as in the case of *dhuA*, where a REP sequence overlaps with the *hisI* promoter and *uvrD*, where a REP sequence is located between the -10 and the initial ATG of the coding sequence. It is also interesting that REP-like sequences have been found in stable RNA species, such as the 16S and 23S rRNA and also in the *ile* tRNA (Table IV).

Stern et al. (1984), have proposed several possible roles for REP sequences. One of them is that REP sequences may simply have a passive role where the two inverted sequences may serve as a block to 3' → 5' exonucleolytic activity at the mRNA level. Alternatively, REP sequences may affect gene expression by altering mRNA stability by serving as a

TABLE IV

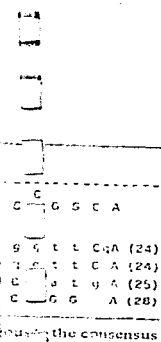
REP-like sequences found as part of genes that code for stable RNA molecules

Name	Sequence
Consensus	G T A G C C C T G A T A A G A C G T 0 to 4 nt G C G T C G C A T C A G G C A
<i>B. subtilis</i> tRNA <i>ile</i>	G C A C C C C T G A T A A G C G T g A C G T C G g T g g t t CgA (24)
<i>E. coli</i> tRNA <i>ile</i>	G C A G C C C T G A T A A G G G T A g G T C G g T g g t t C A (24)
<i>E. coli</i> 16S RNA	G T A G C C C G G A T t G G a G T ctgca A C G T C G AcT C C a t g A (25)
<i>E. coli</i> 23S RNA	G T A a C C C T G C g A A G G C G T gt G C G T GaggCA T C C G A (28)

At the right margin, in parentheses, are specified the numbers of bases (in capital letters), out of 33, that are homologous to the consensus REP sequence. Non homologous bases are in lower-case letters.

This effect seems to be located downstream (see pBB5 vs.

GenBank database number of reported and the number of sequences are presented. Information was obtained from REP sequences (Tables II and III), the consensus REP sequence size, 0-4 or overlapping units reported, these and overlapping *dhcA*, where a *hisJ* promoter and located between the sequence. It has been proposed that the 16S and (Table IV). Several possible role where the mRNA level. By serving as a



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 AATATTC-GCA-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.

Worcel 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.