



CENTRO DE INVESTIGACION SOBRE FIJACION DE NITROGENO



UNIDAD ACADEMICA DE LOS CICLOS PROFESIONAL Y DE  
POSGRADO DEL COLEGIO DE CIENCIAS Y HUMANIDADES

CENTRO DE INVESTIGACION SOBRE FIJACION DE  
NITROGENO, U.N.A.M.

ESTUDIOS SOBRE LA REGULACION DE LA EXPRESION DEL  
OPERON glnALG DE Escherichia coli

TESIS QUE PARA OBTENER EL GRADO DE LICENCIADO EN  
INVESTIGACION BIOMEDICA BASICA PRESENTA

JUAN MIRANDA RIOS

Cuernavaca, Mor., agosto de 1986.

20298



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.

A Angela y Mónica

A Araceli

A mi abue,tía Teresa, tía

Bertha, tío Raúl y primos

A los Cova (Ale, Alejandro,

Martha, Irene, Mario, Paty,

Julio, Rebeca y Jaqueline)

A los cefinitas

## A G R A D E C I M I E N T O S

Este trabajo se realizó en el laboratorio de la Dra. Alejandra Covarrubias del Centro de Investigación sobre Fijación de Nitrógeno (U.N.A.M.). Agradezco a la Dra. Covarrubias sus enseñanzas y su constante preocupación por mi superación académica.

Agradezco al Dr. Fernando Bastarrachea, a la Dra. Guadalupe Espín, al M. en I.B.B. Alejandro Garcíarrubio y al M. en I.B.B. Luis Servín la revisión crítica y su discusión al presente trabajo.

Agradezco al Dr. Ray Sánchez-Pescador por haber obtenido parte de la secuencia presentada aquí.

Agradezco a CONACyT la beca que me otorgó para la realización de esta tesis.



Las estrellas se apagan- dijo el Hermano Gris, olfateando el viento del alba-. ¿Dónde dormiremos hoy? Porque, desde hoy, seguiremos nuevas pistas.

El libro de las tierras vírgenes. R. Kipling.

## R E S U M E N

Se presenta la secuencia nucleotídica de los genes glnL y glnG de E. coli. Se comparó con la reportada de estos mismos genes de K. pneumoniae, observándose que están muy conservadas (homología mayor al 90% a nivel de secuencia de aminoácidos) en estas dos especies.

En el extremo  $-NH_2$  terminal del producto de glnL aparece una región homóloga a la que se ha implicado en la interacción proteínas-DNA. Su extremo  $-COOH$  terminal es homólogo al correspondiente de otras proteínas regulatorias (los productos de los genes dctB, envZ, phoR, cpxA, virA y cheA).

El producto del gene glnG se dividió en 3 dominios funcionales: B, D y E. El dominio B es homólogo a varias proteínas regulatorias, algunas de las cuales regulan la expresión de genes a nivel de transcripción (los productos de los genes ompR, dye, nusA, phoB, spo0A, spo0F, cheB, cheY, dctD y virG). Los dominios D y E son homólogos al extremo  $-COOH$  terminal de los productos de los genes nifA de R. meliloti y K. pneumoniae. En el dominio E se encontró el motivo estructural característico de las proteínas que se unen a DNA.

Se bosqueja un modelo de interacción de estas proteínas regulatorias, en el cual el  $NH_2$  terminal de las proteínas homólogas a glnL capta los estímulos ambientales, transmitiendo esta información al extremo  $COOH$  terminal. Este interactúa con el  $NH_2$  terminal de las proteínas homólogas a glnG. El  $COOH$  terminal de estas últimas proteínas sería la región efectora de las mismas, involucrada en la activación de la transcripción.

## I N T R O D U C C I O N

Las bacterias cuentan con complejos sistemas regulatorios que les permiten adaptarse eficientemente a las condiciones de un medio ambiente cambiante. Uno de tales sistemas regulatorios controla la expresión de las enzimas involucradas en la asimilación de nitrógeno en estos organismos. A este sistema se le conoce como sistema ntr (nitrogen regulation) (1).

En Escherichia coli, Klebsiella pneumoniae, Klebsiella aerogenes y Salmonella typhimurium los productos de los genes glnF, glnL y glnG (o sus equivalentes ntrA, ntrB y ntrC) son los encargados de ejercer tal control (2,3). Entre los genes regulados por este sistema se encuentran: glnA, que es el gene estructural de la enzima glutamina sintetasa (GS: L-glutamato, amonio ligasa, EC 6.3.1.2), genes que codifican para enzimas de las vías degradativas de aminoácidos tales como histidina (hut), arginina (arg) y prolina (put) así como genes que codifican para permeasas de aminoácidos. Los genes nif de K. pneumoniae también están bajo el control del sistema ntr.

Los genes glnL y glnG están contiguos al gene glnA formando el operón glnALG (2,4). Este operón se ha considerado como complejo ya que posee, al menos, un promotor interno (pglnL); este promotor permite la expresión de los genes glnL y glnG separadamente del gene glnA bajo ciertas condiciones metabólicas (5,6,7).

El gene glnF se identificó primero en S. typhimurium (8). No está ligado a glnA; sin embargo, mutaciones en este gene resultan

en niveles bajos constitutivos de GS (20 nanomolas de producto/min/mg de proteína) que convierten a la bacteria en auxótrofa de glutamina (8,9,10). Además, estas mutantes son incapaces de activar a otros genes controlados por la disponibilidad de nitrógeno. El producto del gene glnF junto con el producto del gene glnG actúan como activadores de la transcripción de los genes controlados por nitrógeno en condiciones en que la concentración de amonio está limitada. Actualmente se piensa que el producto del gene glnF cumple su función de regulador positivo al actuar como una subunidad sigma de la RNA polimerasa; esta subunidad se usaría para activar específicamente la expresión de operones regulados por nitrógeno (11,12).

El producto del gene glnG ha resultado ser una proteína bifuncional, ya que además de ser activador, también media la represión de la transcripción del gene glnA en condiciones de exceso de nitrógeno (9) o en ausencia del producto del gene glnF (13). Además, reprime la transcripción del promotor p<sub>glnL</sub> tanto en exceso como en limitación de nitrógeno y en ausencia o presencia del producto del gene glnF (7,14).

El producto del gene glnG es un polipéptido de 55 kd (4), capaz de unirse específicamente al DNA (16,17,18). El dominio -NH<sub>2</sub> terminal de esta proteína parece estar involucrado en la regulación positiva de la transcripción del gene glnA, en tanto que su dominio -COOH terminal parece mediar el control negativo (15). Al examinar la secuencia de DNA de varios promotores en los que el producto del gene glnG actúa como represor, se encontró una región que comparte homología entre los varios promotores analizados y cuya secuencia consenso es: 5' TGCACC<sup>AAAA</sup><sub>TTTT</sub>GGTGCA 3' (16,19,20).

El producto del gene glnL es un polipéptido con un peso molecular aproximado de 36 kd (4). Es, como el producto del gene glnG, una proteína regulatoria bifuncional, cuyo dominio -NH<sub>2</sub> terminal

está involucrado en el control negativo de la expresión de glnA y el dominio -COOH terminal parece actuar en la regulación positiva de otros sistemas del metabolismo del nitrógeno (2,15). No se conoce con precisión su función. Mutaciones en glnL producen una gran variedad de fenotipos: GlnR, síntesis baja constitutiva de GS; GlnC, síntesis alta constitutiva de GS; además, suprimen mutaciones en glnD, glnB y glnF (2,17,21).

Los productos de los genes glnD y glnB son componentes del sistema de adenilación de GS: codifican para la uridililtransferasa (UTasa) y para la proteína P<sub>II</sub>, respectivamente. La adenilación y desadenilación de GS es llevada a cabo por la enzima adenililtransferasa (ATasa). P<sub>II</sub> estimula la actividad adeniladora de la ATasa. Cuando la proteína P<sub>II</sub> es uridilada por la UTasa (convirtiéndola en P<sub>II</sub>-UMP) estimula la actividad desadeniladora. A su vez, la actividad de la UTasa es estimulada por 2-cetoglutarato e inhibida por glutamina. En consecuencia, una baja tasa de glutamina a 2-cetoglutarato, indicativa de deficiencia de amonio, convierte a P<sub>II</sub> a su forma uridilada y resulta en la desadenilación de GS. La proteína P<sub>II</sub>, además de favorecer la adenilación de GS, reprime su síntesis y la de otras enzimas controladas por nitrógeno. La proteína P<sub>II</sub> al ser uridilada por la UTasa, es incapaz de reprimir la síntesis de GS y favorece la desadenilación de la misma (2).

Debido a que una mutación en el gene glnL puede suprimir el fenotipo de mutaciones en los genes glnD y glnB (Gln<sup>-</sup> Ntr<sup>-</sup>), se ha postulado que los productos de los genes glnD y glnB comunican información sobre la disponibilidad de nitrógeno al producto del gene glnL. A su vez, el producto de glnL modularía las actividades de represor-activador del producto de glnG (22,23). En este contexto, se ha sugerido que el producto de glnL puede formar un complejo

con el producto de glnG bajo ciertas condiciones metabólicas (4,17,  
24).

### Un posible modelo de regulación del operón glnALG

El operón complejo glnALG comprende una región contigua de 4500 pares de bases. Está localizado en el minuto 86 del cromosoma de E. coli. Los tres genes (glnA, glnL y glnG) son transcritos en sentido inverso a las manecillas del reloj, del gene glnA hacia el gene glnG (2). Existen 2 promotores al inicio del gene glnA (llamados p<sub>glnA1</sub> y p<sub>glnA2</sub>) y uno más entre glnA y glnL (llamado p<sub>glnL</sub>) (Figura 1). Además, existe un terminador independiente de factor rho al final del gene glnA (5).

Un posible modelo de regulación de la expresión del operón glnALG de E. coli es el siguiente:

Cuando la bacteria está creciendo en un medio limitado de nitrógeno, la relación 2-ceto-glutarato/glutamina aumenta, indicando una deficiencia de amonio. Este cambio en la relación es "sentido" por la UTasa, la cual uridila a la proteína  $P_{II}$ . De alguna forma, estas proteínas transmiten esa información al producto de glnL, el cual convierte al producto de glnG a su forma activadora. De esta manera, el producto de glnG junto con el producto de glnF (que es una subunidad sigma de la RNA polimerasa) y la RNA polimerasa iniciarían la transcripción a partir de p<sub>glnA2</sub>. Este transcrito terminaría con alta frecuencia en el terminador que se encuentra en la región intercistronica entre glnA y glnL. Muy poco de la transcripción continuaría a través de glnL y glnG, aunque ésta sería suficiente para mantener niveles adecuados del producto de glnG necesarios para la activación de p<sub>glnA2</sub> y de otros promotores regulados por nitrógeno. En esta condición, tanto p<sub>glnA1</sub> como p<sub>glnL</sub> estarían reprimidos por el

producto de glnG.

En crecimiento en exceso de nitrógeno, el producto de glnL, en base a la información que sobre este estado metabólico le transmiten la UTasa y la proteína P<sub>II</sub>, convierte al producto de glnG a su forma inactiva. Esto produce una menor activación en pglnA2, con la consiguiente disminución tanto del transcrito que termina en el terminador que está al final de glnA como del que termina al final del operón. En esta condición, pglnA1 y pglnL estarían menos reprimidos. Los transcritos así originados servirían para mantener niveles suficientes de los productos de glnL y glnG que le permitan a la bacteria reaccionar ante una posterior limitación de nitrógeno.



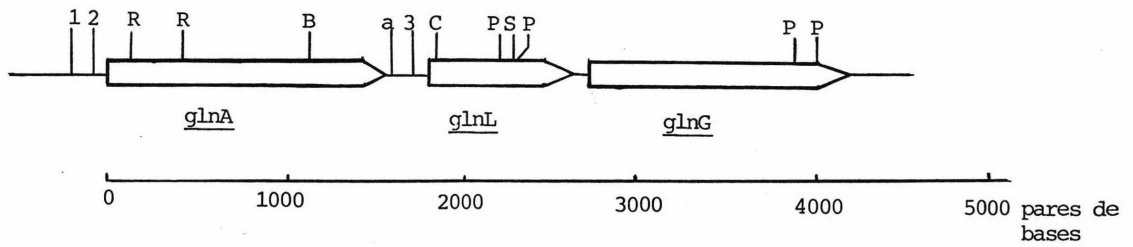


FIGURA 1: Mapa físico del operón *glnALG* de *E. coli*.  
 1= *pglnA1*; 2= *pglnA2*; 3= *pglnL*  
 a= terminador independiente de factor rho.  
 Sitios de restricción: R= *Eco-R1*; B= *Bam-H1*;  
 C= *Cla-1*; S= *Sal-1*; P=*Pvu-II*.

## O B J E T I V O S

El objetivo de este trabajo es estudiar a nivel molecular a los genes glnL y glnG de Escherichia coli y su participación como reguladores del sistema ntr. Para tal efecto y como un primer paso hacia la identificación de los dominios funcionales de estos dos reguladores, se determinó la secuencia nucleotídica de los genes glnL y glnG, analizándose y comparándose con la secuencia reportada de estos mismos genes de K. pneumoniae y de otras proteínas regulatorias pleiotrópicas de procariotes.

## M A T E R I A L Y M E T O D O S

### Cepas bacterianas y plásmidos

Las cepas bacterianas y los plásmidos usados en este trabajo se enlistan en la tabla I.

### Bioquímica del DNA

La purificación del DNA, la electroforesis del mismo, la digestión con enzimas de restricción y su clonación en los plásmidos M13 mp18 y mp19 se llevaron a cabo de acuerdo a (25). La secuencia se realizó usando el método de Sanger y Coulson (26). En el apéndice se amplía la información sobre los vectores y el método de secuenciación usados.

### Estrategia para obtener la secuencia nucleotídica de los genes glnL y glnG de E. coli.

Los mapas físicos de los plásmidos pACR3 y pACR34 se usaron para diseñar la estrategia de secuencia (Figura 2).

El plásmido pACR3 contiene al gene glnL excepto la región que codifica para los primeros 23 aminoácidos. En los plásmidos M13 mp18 y mp19 se clonó el fragmento Sal-I - Eco-R1 de 480 pares de bases. Este fragmento se secuenció en ambas orientaciones. Con objeto de resolver las ambigüedades de esta secuencia, se digirió el mismo fragmento con la enzima Hpa-II. Los fragmentos resultantes

<u>Cepas</u>	<u>Genotipo</u>	<u>Referencia</u>
<u>E. coli</u> JM101	( <u>lac pro</u> ) <u>thi str supE endA</u> <u>sbcB hsdR F'</u> ( <u>traD36 proAB</u> ) <u>lacI<sup>q</sup> ZdelM15</u>	(49)

<u>Plásmidos</u>	<u>Fenotipo Relevante</u>	<u>Referencia</u>
M13 mp18, mp19	rep <sub>M13</sub>	(48)
pACR3	rep <sub>ColE1</sub> Ap <sup>r</sup>	(27)
pACR34	rep <sub>ColE1</sub> Ap <sup>r</sup>	(27)

Tabla I. Cepas bacterianas y plásmidos.

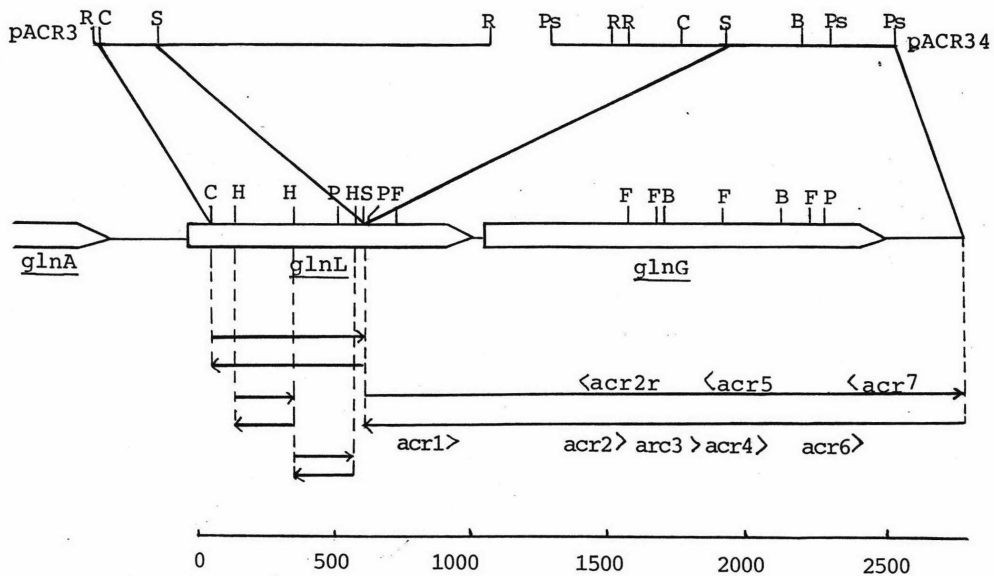


FIGURA 2: Estrategia de secuancia de los genes *glnL* y *glnG* de *E. coli*. La porción de la secuancia nucleotídica obtenida de cada clona de M13 se muestra debajo del mapa por medio de flechas. Las puntas de flecha ( ) indican la posición de los oligonucleótidos sintéticos que se usaron como "primers", junto con su nombre. C= *Cla*-I; H= *Hpa*-II; P= *Pvu*-II; S= *Sal*-I; F= *Hinf*-I; B= *Bal*-I; Ps= *Pst*-I; R= *Eco*-RI.

se clonaron en M13 mp18 y se secuenciaron en ambas orientaciones.

Por otro lado, para obtener la secuencia del extremo 3' del gene glnL y la secuencia total del gene glnG, se clonó en M13 mp18 y mp19 el fragmento Sal-I - Pst-I proveniente de pACR34. Se secuenciaron 400 pares de bases de cada una de las clonas. Se sintetizaron dos oligonucleótidos (acr1: 5' TGATTGATCAGCATTCAGGC 3'; acr7: 5' AGCCGCGCCGCTTCCTGTTTATG 3') los cuales se usaron como "primers" para secuenciar las siguientes 480 y 490 pares de bases, respectivamente. Con la secuencia obtenida, fue posible sintetizar otros dos oligonucleótidos (acr2: 5' GTAATGTTTCAGCTTAACGGC 3'; acr5: 5' TCGAGATTCTGGTGAGTGGC 3') los cuales a su vez permitieron secuenciar las siguientes 240 y 220 pares de bases. Conociendo esta secuencia, se sintetizaron dos oligonucleótidos (acr3: 5' TGGCGCGAATACCATTCGTC 3'; acr2reverso: 5' TCGAGATTCTGGTGAGTGGC 3') con los que se alargó la secuencia en 180 y 480 pares de bases. Se sintetizó el oligonucleótido acr4 (5' AGTGCAGGAAGGTAAGTTCC 3') para secuenciar 320 p.b. más y el acr6 (5' CGACGCTTCTTGCGCAGTGG 3') para secuenciar lo que restaba del fragmento.

## R E S U L T A D O S

Se obtuvo una secuencia de 2602 pares de bases que se extiende del sitio Cla-I en el extremo 5' de glnL hasta el inicio de las colas A:T que se usaron para clonar este fragmento en pBR327 y que se encuentran 205 pares de bases después del final del gene glnG Figura 3.

Los datos revelan dos fases abiertas de lectura. La primera se extiende 974 pares de bases a la derecha del sitio Cla-I, la cual se identificó como la secuencia del gene glnL. La segunda se extiende por 1404 pares de bases y corresponde al gene glnG.

### El producto del gene glnL de E. coli

El gene glnL codifica para un polipéptido de 349 aminoácidos con un peso molecular de 38,647 d calculado a partir de la secuencia de nucleótidos. El peso molecular es muy semejante al reportado de 36,000 (4).

### Composición de aminoácidos del producto del gene glnL

Al comparar la composición de aminoácidos del producto del gene glnL con la de una proteína promedio calculada en base a los datos de secuencia de 314 proteínas de E. coli, la única característica sobresaliente es el contenido relativamente alto de residuos hidrofóbicos (29.5%) comparado con el de una proteína promedio (20.2%) (53) (Tabla II).

1 ATGGCAACAGGCACACTGCCCGATGCTGGGCAGATCCTTAACTCTTTAATCAACAGTATCTTACTGGTCGAT  
 \*\*\*\*\* \* \*\*\*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\*  
 1 ATGGCAACAGGCACAGCCCGATGCTGGGCAGATCCTCAACTCGCTGATTAACAGTATTTTGTTAATCGAT

73 GACGACCTGGCGGTACATTACGGGAACCCGGCGGGCAGCAGCTGCTGGCCCAAGCTCCGCTAAATTATTC  
 \*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\* \*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\*  
 73 GACAACCTGGCGATCCATTACGCCAACCCGCGCGGCAACAACCTGCTGGCCCAAGCTCCCGCAAATTGTTT

145 GGCACGCCGCTGCCGGAGCTGTTGAGCTATTTCTCGCTGAATATCGGCCTCATGCAAGAGAGCCTCGCTGCG  
 \*\* \*\* \*\*\* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \*\*\*\*\* \*\* \*\*\*\*\* \* \* \* \* \*  
 145 GGTACACCGTTACCGGAAGCTGTTGAGCTACTTCTCATTAAATATCGAGCTGATGCAAGAAAGTCTGGAGGCC

217 GGTACAGGGTTTACCGATAACGAAGTGACGCTGGTTATTGATGGCCGCTCCACATCCTTTCCCTGACGGCG  
 \*\* \*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\* \*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\*  
 217 GGGCAAGGTTTACCGATAACGAAGTGACGCTGGTCATCGACGGCGGCTCGCATATCCTTTCTGTGACGGCC

289 CAACGCCCTGCCTGAAGGCTATATCCTGCTGGAATGGCCGGGATGGATAATCAGCGTCCGCTGAGCCAGGAG  
 \*\* \*\* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \* \* \*  
 289 CAGCGTATGCCGGACGGCATGATCCTGCTGGAGATGGCTCCGATGGATAACCAGCGCCGCTTAAGTCAGGAA

361 CAGCTACAGCACGGCAGCAGATTGCCGCACGTGACTTAGTGCCTGGTCTGGCCACAGATTAAAAATCCG  
 \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\* \*\*\*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\*  
 361 CAGCTACAGCACGCCAGCAGGTTGCTGCCCGTGATTTAGTGCCTGGCCACATGAGATTAAAAATCCG

433 CTGGGGGGCCTGGCGGGCGGGCGCAGCTGCTGAGCAAAGCGCTGCCCGACCCTGCGCTAATGGAAATATACC  
 \*\* \*\*\*\*\* \* \*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \*\* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\*  
 433 CTTGGCGGTTTACGTGGCGGGCGCAGCTGCTCAGCAAAGCGTTACCTGACCCATCACTACTCGAATATACC

505 AAAGTCATTATTGAACAGGCCGACCGGTTAAGAAATCTGGTCGACCGTCTACTGGGACCACAGCATCCCGGA  
 \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\*  
 505 AAAGTGATTATCGAACAGGCCGGACCGGCTGCGAAATCTGGTCGACCGTCTGTTGGGGCCGAGCTGCCCGG

577 ATGCACGTAACAGAAAGCATCCACAAGGTCGCTGAGCGAGTGGTGAAGCTGGTGTGATGGAACTACCGGAT  
 \* \*\* \*\*\* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*  
 577 ACGCGCGTTACCGAAAGTATTCACAAGGTGCTGAACGCGTGGTAACGCTGGTGTGATGGAACTGCCGGAC

649 AACGTCAAGCTGGTGGCGACTATGACCCAGCCTGCCAGAGCTCCCGCACGACCCCGATCAAATTGAACAA  
 \*\*\*\*\* \* \*\* \* \* \* \* \* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*  
 649 AACGTGCGGTTGATTGCTGATTACGATCCAGCCTACCGGAAGTGGCGCACGACCCCGATCAAATTGAACAG

721 GTGCTGCTGAACATCGTACGTAACCGGCTGCAGGGCGCTGGGACCGGAGGGCGGAGAGATCACTCTGCGCACC  
 \*\* \*\*\*\*\* \* \* \* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\* \* \* \*  
 721 GTCTTGCTGAATATTGTGCGCAATGCGCTACAGGGCGCTGGGGCCGGAAAGGCGGTGAATCATTCTGCGTACC

793 CGGACGGCCTTCCAGTTAACCCCTGCACGGCGTGGCTACCGCCTCGCTGGCGGTATAGCGTTGAAGATAAC  
 \*\* \*\* \* \* \* \* \* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\*  
 793 CGCACCGCGTTTCAACTGACCTTACACGGCGAGCCCTACCGGCTGGCGGGCGGATTGATGTGGAAGATAAC

865 GGGCCGGGAATCCCGTCTCACTTACAGGATACGCTGTTTTATCCCATGGTGAAGTGGACGGCAAGGCGGTACC  
 \*\* \*\*\*\*\* \* \* \* \* \* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\*  
 865 GGGCCGGGCATTCCCGCTCATTTGCAGGATACGCTGTTTTACCGGATGGTCAGCGGCCCGGAAGGTGGCACC

937 GGAATCGGGCTGTCTATCGCCCGCAGCCTTATCGACCAGCACTCCCGCAAATTTGAATTTACCAGTTGCCCA  
 \*\* \*\* \* \* \* \* \* \*\*\*\*\* \* \* \* \* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*  
 937 GGGCTTGGCTTATCCATCGCTCGTAATTTGATTGATCAGCATTTCAGGCAAATTTGAATTTACCAGTTGCCCA



009 GGGCATACCGAATTCTCGGTGTACCTGCCTATTCGGAAAGTAGAGTG TTTATGCAACGAGGGATAGCAT  
 \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*  
 009 GGGCATACCGAGTTCCTCGGTTACCTGCCTATCAGGAAATAAGAGGTGACGTTTATGCAACGAGGGATAGTCT

081 GGATCGTTGAC  
 \*\* \* \* \* \*  
 081 GGGTAGTCGAT



1 ATGCAACGAGGGATAGCATGGATCGTTGACGACGATAGCTCCATCCGTTGGGTGCTTGAACGCGCGCTCACC  
\*\*\*\*\*  
1 ATGCAACGAGGGATAGTCTGGGTAGTTCGATGACGATAGTTCCATCCGTTGGGTGCTTGAACGTCGCGCTCGCT  
  
73 GGGGCAGGCTTGAGTTGTACAACGTTTGAAGCGGCAATGAGGTGCTTGTATGCCCTCACGACCAAACCCCG  
\*\*\*\*\*  
73 GGGGCAGGTTTAACTGTACGACGTTTGAAGACGGCGCAGAAGTGTCTGGAGGGCGCTGGCGAGCAAACCCCG  
  
145 GATGTACTGCTGTCTGATATCCGCATGCCGGGGATGGATGGGCTGGCGCTGCTAAAGCAAATTAACAGCGT  
\*\*\*\*\*  
145 GATGTGCTGCTTTCAGATATCCGATGCCGGGAATGGACGGGCTGGCGCTGCTCAAGCAGATTAACAGCGC  
  
217 CATCCGATGCTCCCGTCATCATAATGACGGCGCATTCCGATCTTGACGCGGGCGTCAAGCGCCTATCAGCAG  
\*\*\*\*\*  
217 CATCCAATGCTTCCCGTCATCATTATGACCGCACATTCCGATCTGGATGCTGCCGTCAGCGCCTATCAACAA  
  
289 GGCGCTTTTGATTATCTGCCGAAACCGTTTGTATATTGATGAAGCGGTGGCGCTGGTCCGACCGTGGCGATTAGC  
\*\*\*\*\*  
289 GGGGCGTTTGATTATCTGCCCAAACCGTTTGTATATCGACGAAGCAGTGGCGCTGGTTGAGCGCGCTATCAGT  
  
361 CATTACCAGGAACAGCAGCAGCCGCGTAATGGCGCGATTAATAGCCCGACGGCCGATATTATCGCCGAAGCG  
\*\*\*\*\*  
361 CATTACCAGGAACAGCAGCAGCCGCGTAATGTTTCAGCTTAACGGCCCAACGACCGATATCATCGCGAAG  
  
433 CCGGCCATGCAGGACGTTTTTTCGATTATCGGCCGCGCTGTCCGCTCATCAATCAGCGTACTGATCAACGGG  
\*\*\*\*\*  
433 CCAGCCATGCAGGACGTTTCCGATTATCGGTCGGCTTTCGCGTCTTCTATTAGCGTGTGATTAACGGC  
  
505 GAATCCGGTACCGGTAAGAGCTGGTTGCCCATGCTCTCCATCGGCATAGTCCGCGCGCCAAGGGCGCCGTTT  
\*\*\*\*\*  
505 GAATCCGGCACCGGTAAGAACTGGTCCGTCATGCCCTGCATCGCCACAGTCCGCGCGCCAAGGGCGCCGTTT  
  
577 ATTGGCTAAATATGGCAGCAATTCCTAAAGATTTAATCGAATCAGAGCTATTTGGCCACGAGAAAGGGGCC  
\*\*\*\*\*  
577 ATCGCGCTGAATATGGCAGCTATCCCAAAGATTTGATCGAATCAGAACTGTTTGGCCACGAGAAAGGGCGC  
  
649 TTTACCGGCGCCAATACCGTGCCTCAGGGACGCTTTGAGCAGGCAGACGGCGGCACGCTATTCCTCGACGAG  
\*\*\*\*\*  
649 TTTACTGGCGGAATACCATTCGTCAGGGCGCTTTGAACAGGCCGATGGCGGTACATTATTCCTCGACGAA  
  
721 ATCGGCGATATGCCGCTGGACGTACAGACTCGCCTGCTGGCGCTCCTCGCTGATGGCCAGTTTTATCGCGTC  
\*\*\*\*\*  
721 ATTGGTGATATGCCGCTGGATGTGCAGACGCGTGTGCTGGCGCTGCTGGCAGACCGTCAGTTTTACCGCGTT  
  
793 GCGGGCTATCCGCGGTGAAAGTGGATGTGGGCATCATTGCCGCCACTCACCAGAACCTTGAGCTGGCGGTC  
\*\*\*\*\*  
793 GCGGGCTATGGCGCGGTGAAAGTGGATGTGGGGATTATCGCTGCCACTCACCAGAATCTCGAACAGCGAGTG  
  
865 CAGGAAGGAAAATTCGCGAGGATTTATTCCACCGCCTGAACGTGATTCGCGTCCATCTGCCGCCGCTGGCG  
\*\*\*\*\*  
865 CAGGAAGGTAAGTTCGCTGAGGATCTGTTCCACCGCCTGAACGTTATCCGCGTTCATCTGCCGCCGCTGGCG  
  
937 GAGCGTCGGGAAGATATTCCCTCCCTGGCGAGGCACCTCCTGCAAAATTGCCGCTCGTGAACGGGGCGTCGAG  
\*\*\*\*\*  
937 GAACGTGGGAAGATATTCCCGCTCTGGCGCGCCATTTTTTACAGGTTGCCGCGCGCGAACTGGGGCGTAGAA  
  
009 GCCAAACAGCTGCATCCGGAGACGGAGATGGCGTTAACCCGACTGGCATGGCCGGGTAACTACGTCAGCTG  
\*\*\*\*\*  
009 GCGAAGTTACTGCATCCGGAAACCGAAGCTGCTCTGACGCGTCTGGCGTGGCCGGGCAACGTGGCCAGCTG  
  
081 GAAAACACCTGTCCCTGGCTGACGGTAATGGCCGCAGGTCAGGAGGTTCTGACGCAAGGATCTTCCCAGCGAG  
\*\*\*\*\*  
081 GAAAACACCTGCCGCTGGCTAACGGTGATGGCCGCCGGGCAAGGAGTGTGATTCAAGATTTGCCCGCGCGAA  
  
153 CTGTTTGAACCGCGATACCGGATAACCCAACGCAAATGCTGCCGGATAGCTGGGCCACGCTGCTGGGACAA  
\*\*\*\*\*

L153 CTGTTTGAAACCGCGATACCGGATAACCCAACGCAAATGCTGCCGGATAGCTGGGCCACGCTGCTGGGACAA  
\*\*\*\*\* \* \*\* \* \*\*\*\* \* \* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\* \* \*\*  
L153 CTGTTTGAATCAACGGTTGCGGAGAGTACTTCGCAAATGCAACCGGACAGCTGGGCGACGCTTCTTGGGCAG  
L225 TGGGCCGATCGCGCTTTGCGTTCCGGTCATCAAAACCTGCTCTCGGAAGCGCAGCCTGAGATGGAGCCGACG  
\*\*\*\*\* \*\* \* \*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\* \*\*\*\*\* \*\*  
L225 TGGGCAGACAGAGCGCTGCGTTCCGGTCATCAAAATCTGCTTTCCGAAGCGCAGCCAGAGCTGGAGCCGACG  
L297 CTGCTGACCACCGCCTTACGCCACACTCAGGGGCATAAACAGAGGCGGCCCGGCTGCTGGGATGGGGTCGT  
\* \*\*\*\*\* \*\*\*\*\* \*\* \*\* \*\* \* \*\*\*\*\* \*\*\*\*\* \*\* \*\*\*\*\* \*\* \*\* \*\*\*\*\* \*\*  
L297 TTA CTGACGACCGCGTTGCGACATACGCAGGGGCATAAACAGGAAGCGGCGCGGCTACTCGGCTGGGGCCCG  
L369 AATACCCTGACCGGTAAGCTAAAAGAGTTGGGGATGGAGTAG  
\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\* \*\*\*\*\*  
L369 AACACCCTGACCGGTAAGTTAAAAGAGCTGGGGATGGAGTGATTCACAGCTTGTGTGTAAAGATTGATTATT

FIGURA 3. Comparación de las secuencias nucleotídicas de los genes glnL y glnG de E. coli (abajo) y K. pneumoniae (arriba). Los asteriscos (\*) indican homología entre ambas secuencias. En la primera página se presenta de secuencia de glnL (las primeras 1047 pares de bases). En las siguientes 2 páginas se presenta la secuencia de glnG (1404 pares de bases). El cuadro  indica los codones de terminación y la  el sitio de unión a ribosomas.

<u>Residuos</u>		producto de <u>glnL</u> %	producto de <u>glnG</u> %	proteína promedio %
alifáticos pequeños	A+G	14.6	17.3	16.9
hidroxílicos	S+T	10.6	9.8	13.1
acídicos	D+E	12.0	12.6	11.6
acídicos+acidamidas	D+N+E+Q	22.6	21.4	19.8
básicos	K+R+H	12.0	14.5	13.5
hidrofóbicos	L+V+I+M	29.5	26.3	20.2
aromáticos	F+Y+W	4.3	5.5	8.3
cargados	D+E+K+R+H	24.1	27.1	25.1

Tabla II: composición de aminoácidos de los productos de los genes glnL y glnG de E. coli comparados con la respectiva de una proteína promedio de E. coli. (53). Los aminoácidos están designados usando el código de una letra.

### Distribución de aminoácidos en el producto del gene *glnL*

Se analizó la distribución de aminoácidos cargados en el producto del gene *glnL*. La proteína tiene una carga neta de -1. Los residuos ácidos y básicos están uniformemente distribuidos a lo largo de la proteína, a excepción de la región entre los residuos 54-86 en la que existen 7 cargas negativas juntas, sin ninguna carga positiva entre ellas. En el producto del gene *glnL* no hay regiones predominantemente hidrofóbicas, por lo que es muy probable que sea una proteína soluble. (Figura 4).

### Frecuencia de uso de codones en el gene *glnL*

La frecuencia de uso de codones "óptimos" en los genes de *E. coli* está relacionada con los niveles de producción de las proteínas codificadas por dichos genes (51). En el caso de genes altamente expresados (por ejemplo *tufA* y *rplA*) casi todo el gene está ocupado por codones óptimos. Sin embargo, en el caso de genes moderadamente expresados (como *trpC* y *trpA*) o débilmente expresados (como *trpR*) la ocurrencia de codones óptimos es menor. La incidencia de 23 codones "no óptimos" en el gene *glnL* es de 26%. Esto está de acuerdo a lo reportado para proteínas regulatorias, en cuyos genes la incidencia es de 24%, mientras que para genes de proteínas no regulatorias es de 12% (45). (Tabla III).

Homología de secuencia entre el producto del gene *glnL* y otras proteínas que se unen a DNA.

Estudios cristalográficos de la interacción del represor y

```
1 MATGTLPDAGQILNSLINSILLVDDDLAVHYANPAAQQLLAQSSRKLFGTPLPELLSYFSLNIGLMQESLAA
*****
1 MATGTQPDAGQILNSLINSILLIDDNLRIHYANPAAQQLLAQSSRKLFGTPLPELLSYFSLNIELMQESLEA
*****
73 GGGFTDNEVTLVIDGRSHILSLTAQRLPEGYILLEMAMPDNQRRLSQEQLDHAQQIAARDLVRGLAHEIKNP
*****
73 GGGFTDNEVTLVIDGRSHILSVTAQRMPDGMILLEMAMPDNQRRLSQEQLDHAQQVAARDLVRGLAHEIKNP
*****
145 LGGLRGAAQLLSKALPDPALMEYTKVIEQADRLRNLDVDRLLGPOHPGMHVYTESIHKVAERVVKLVSMELPD
*****
145 LGGLRGAAQLLSKALPDPSLLEYTKVIEQADRLRNLDVDRLLGPOHPGTRVYTESIHKVAERVVTLVSMELPD
*****
217 NVKLVRDYDPSLPELPHDPDQIEQVLLNIVRNALQALQPEGGEILRTRTAFQLTLHGVRYLAAARIDVEDN
** *
217 NVRLIRDYDPSLPELAHDPDQIEQVLLNIVRNALQALQPEGGEILRTRTAFQLTLHGGERYLAAARIDVEDN
*****
289 GPGIPSHLQDTLFYPMVSGREGGTGLGLSIARSLIDQHSKIEFTSWPGHTEFSVYLPARK
*****
289 GPGIPPHLQDTLFYPMVSGREGGTGLGLSIARNLIDQHSKIEFTSWPGHTEFSVYLPARK
*****
```

FIGURA 4. Comparación de las secuencias de aminoácidos deducidas de las secuencias nucleotídicas de los productos de glnL de E. coli (abajo) y K. pneumoniae (arriba). Los asteriscos (\*) indican homología entre ambas secuencias. Los aminoácidos se designan de acuerdo al código de una letra.

5/UUU/F	1/UCU/S	1/UAU/Y	0/UGU/C
2/UUC/F	2/UCC/S	6/UAC/Y	0/UGC/C
9/UUA/L	3/UCA/S	1/UAA/oc	0/UGA/op
8/UUG/L	4/UCG/S	0/UAG/am	1/UGG/W
3/CUU/L	4/CCU/P	6/CAU/H	7/CGU/R
4/CUC/L	3/CCC/P	4/CAC/H	11/CGC/R
4/CUA/L	2/CCA/P	7/CAA/Q	1/CGA/R
25/CUG/L	13/CCG/P	16/CAG/Q	5/CGG/R
13/AUU/I	0/ACU/T	6/AAU/N	5/AGU/S
11/AUC/I	9/ACC/T	8/AAC/N	5/AGC/S
0/AUA/I	2/ACA/T	7/AAA/K	0/AGA/R
8/AUG/M	6/ACG/T	0/AAG/K	1/AGG/R
3/GUU/V	5/GCU/A	12/GAU/D	6/GGU/G
4/GUC/V	6/GCC/A	8/GAC/D	12/GGC/G
10/GUG/V	12/GCG/A	6/GAG/E	8/GGG/G

Tabla III

Tabla de uso de codones del gene glnL de E. coli. Los aminoácidos se designan empleando el código de una letra.

de la proteína Cro de lambda y de la proteína CAP de E. coli con el DNA, indican que estas proteínas reconocen al DNA de manera similar: todas presentan un motivo estructural que consiste en dos alfa hélices consecutivas (alfa-2 y alfa-3) separadas por un giro. La hélice alfa-3 encaja en el surco mayor de la doble hélice del DNA e interacciona con los pares de bases presentes en el sitio de unión; la hélice alfa-2 interactúa con el esqueleto de azúcar-fosfato del DNA (Figura 5). En la secuencia de aminoácidos del producto del gene glnL se encontró una región entre los residuos 67 y 86 que presenta una extensa homología a los dominios que se presume interaccionan con el DNA de manera específica (Figura 6).

La secuencia de aminoácidos del producto del gene glnL se asemeja a la secuencia de otras proteínas que se unen al DNA en lo siguiente:

1) Presenta un aminoácido hidrofóbico en las posiciones 4 y 10 del motivo estructural (corresponde a los residuos 70 y 76).

2) presenta una glicina en la posición 9 (residuo 75) que se cree juega un papel muy importante en la formación del giro entre las dos hélices.

3) presenta una valina en la posición 15 (residuo 81).

El producto del gene glnL presenta, además, dos diferencias notables: en la posición 5 (residuo 71), en lugar de tener una alanina, tiene un ácido glutámico; en la posición 8 (residuo 74) tiene una glutamina en lugar de un residuo hidrofóbico.

Homología entre las secuencias nucleotídicas de los genes glnL de E. coli y K. pneumoniae

Las secuencias nucleotídicas de los genes glnL de E. coli



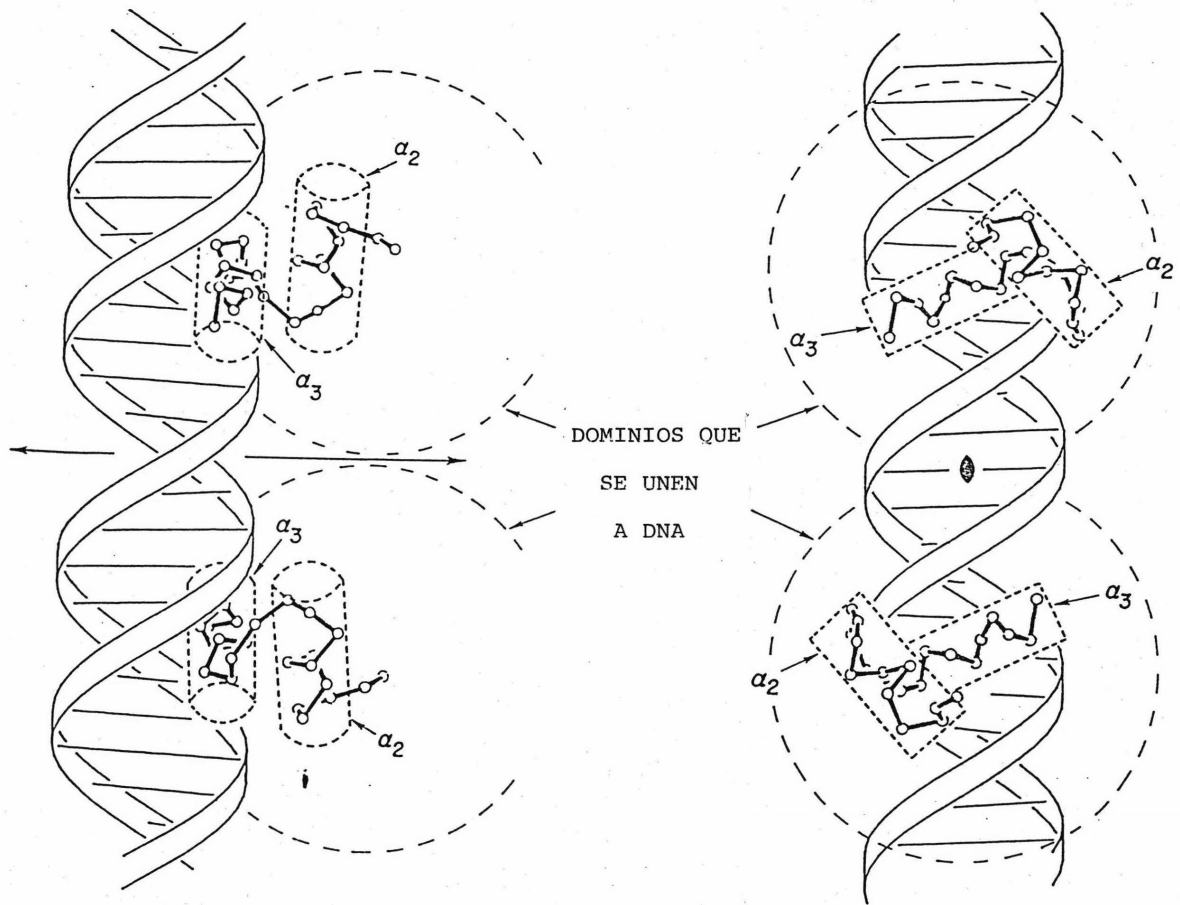


Figura 5. Modelo de la interacción entre proteínas regulatorias y DNA.

A la izquierda se presenta una vista de lado y a la derecha una vista de frente (43).

y de K. pneumoniae (28) (Figura 3) presentan una homología del 81%.

La homología entre los aminoácidos deducidos de la secuencia nucleotídica de glnL en estos dos organismos es del 93%.

Homología entre el dominio -COOH terminal del producto del gene glnL con el dominio correspondiente de otras proteínas regulatorias.

Al analizar las secuencias de aminoácidos deducidos de la secuencia nucleotídica del gene glnL, se observa una alta conservación de los dominios -COOH terminal de los productos de los genes glnL, dctB (gene regulatorio del transporte de C-4 dicarboxilatos en Rhizobium meliloti), envZ (gene regulatorio de la síntesis de porinas en E. coli), phoR (gene regulatorio del metabolismo de fosfato), cpxA (gene regulatorio de la transferencia de plásmidos F), virA (gene regulatorio contenido en el plásmido Ti de Agrobacterium tumefaciens y que está involucrado en la virulencia de esta bacteria) y cheA (gene regulatorio de la síntesis de proteínas quimiotácticas) (52).

El producto del gene glnG de E. coli

El gene glnG codifica para un polipéptido de 468 aminoácidos con un peso molecular calculado de 52,205 d (el peso molecular reportado es de 50,000 d (4)).

Composición de aminoácidos del producto del gene glnG

Al igual que para el producto del gene glnL, el único dato

sobresaliente acerca de la composición de aminoácidos del producto del gene glnG es el contenido relativamente alto de residuos hidrofóbicos (26.3%) comparado con el de una proteína "promedio" (20.2%) (53) (Tabla II).

#### Distribución de aminoácidos en el producto del gene glnG

El producto del gene glnG es una proteína básica con una carga neta de +10.

La distribución de los residuos cargados es uniforme a lo largo de la proteína (Figura 7), aún cuando existen 5 zonas en que se concentran cargas de un signo:

entre los residuos 144-227 existen 16 residuos básicos y 7 ácidos (carga neta= +9);

entre los residuos 229-247 se encuentran 6 residuos ácidos (carga neta= -6);

entre los residuos 293-324 existen 12 residuos básicos y 6 ácidos (carga neta= +6);

entre los residuos 376-401 se localiza un residuo básico y 6 ácidos (carga neta= -5);

entre los residuos 439-469 se encuentran 9 residuos básicos y 3 ácidos (carga neta= +6).

No se encontró ninguna región predominantemente hidrofóbica en el producto del gene glnG, por lo que se sugiere que es una proteína soluble.

```

1 MQRGIAWIVDDSSIRWVLERALTGAGLSCTTFESGNEVLDALTTKTPDVLLSDIRMPGMDGLALLKQIKQR
***** * *****
1 MQRGIVWVVDSSIRWVLERALAGAGLTCTTFENGAEVLEALASKTPDVLLSDIRMPGMDGLALLKQIKQR

73 HPNLPVIIMTAHSDLDAAVSAYQQGAFDYLPKPFIDIEAVALVDRAISHYQEQQQPRNAPINSPTADIIGEA
***** ***** * **
73 HPNLPVIIMTAHSDLDAAVSAYQQGAFDYLPKPFIDIEAVALVERAISHYQEQQQPRNVQLNGPTTDIIA K

145 FAMQDVFRIIGRLSRSSISVLINGESGTOKELVAHALHRHSPRAKAPFIALNMAAIPKDLIESELFGEKGA
*****
145 FAMQDVFRIIGRLSRSSISVLINGESGTOKELVAHALHRHSPRAKAPFIALNMAAIPKDLIESELFGEKGA

217 FTGANTVRQGRFEGADGGTLFLDEIGDMPLDVQTRLLRVLADGGQFYRVGGYAPVKVDVRIIAATHQNLRLV
***** ***** **
217 FTGANTIRQGRFEGADGGTLFLDEIGDMPLDVQTRLLRVLADGGQFYRVGGYAPVKVDVRIIAATHQNLQRV

289 QEGKFREDFHRLNVIRVHLPPLRERREDIPRLARHFLOIAARELGVEAKQLHPETEMALTRLAWPQNVRL
***** *****
289 QEGKFREDFHRLNVIRVHLPPLRERREDIPRLARHFLOVAARELGVEAKLLHPETEALTRLAWPQNVRL

361 ENTCRWLTVMAAGQEVLTQDLPSSELFETAIPDNPTQMLPDSWATLLGQWADRALRSQHONLLSEAGPEMERT
***** ***** ** *****
361 ENTCRWLTVMAAGQEVLIQDLPGELFESTVAEESTSQMOPDSWATLLAQWADRALRSQHONLLSEAGPELERT

433 LLTTALRHTQGHKQEAARLLGWRNTLTRKLELGE
*****
433 LLTTALRHTQGHKQEAARLLGWRNTLTRKLELGE

```

FIGURA 7. Comparación de las secuencias de aminoácidos deducidas de las secuencias nucleotídicas de los productos de glnG de E. coli (abajo) y K. pneumoniae (arriba). Los asteriscos (\*) indican homología entre ambas secuencias. Los aminoácidos se designan de acuerdo al código de una letra.

### Frecuencia de uso de codones en el gene glnG

La incidencia de uso de 23 codones "no óptimos" en el gene glnG es de 20%, valor semejante al que presentan los genes de otras proteínas regulatorias en los cuales es del 24% (45).

### Homología del producto de glnG con proteínas que se unen a DNA,

La secuencia de aminoácidos del producto de glnG en la región comprendida entre los residuos 444-463 es altamente homóloga a la involucrada en las interacciones específicas entre proteínas y DNA.

Presenta residuos hidrofóbicos en las posiciones 8 (residuo 451), 10 (residuo 453) y 15 (residuo 458); una alanina en la posición 5 (residuo 448) y una glicina en la posición 9 (residuo 452).

### Homología entre las secuencias nucleotídicas de los genes glnG de E. coli y K. pneumoniae

Las secuencias nucleotídicas de los genes glnG de E. coli y K. pneumoniae (29,30) (Figura 3) son homólogas en un 82%, mientras que la homología entre sus aminoácidos deducidos de la secuencia nucleotídica es de 93%.

### Homología entre el producto de glnG y los productos de otros genes regulatorios procarióticos.

En un trabajo anterior (30), el producto de glnG de K. pneumo-

10/UUU/F	2/UCU/S	3/UAU/Y	1/UGU/C
4/UUC/F	5/UCC/S	2/UAC/Y	1/UGC/C
6/UUA/L	3/UCA/S	0/UAA/oc	1/UGA/op
5/UUG/L	2/UCG/S	0/UAG/am	7/UGG/W
8/CUU/L	0/CCU/P	11/CAU/H	15/CGU/R
5/CUC/L	3/CCC/P	4/CAC/H	14/CGC/R
2/CUA/L	6/CCA/P	6/CAA/Q	3/CGA/R
33/CUG/L	13/CCG/P	23/CAG/Q	5/CGG/R
10/AUU/I	3/ACU/T	5/AAU/N	4/AGU/S
13/AUC/I	9/ACC/T	7/AAC/N	4/AGC/S
1/AUA/I	1/ACA/T	10/AAA/K	1/AGA/R
11/AUG/M	13/ACG/T	5/AAG/K	0/AGG/R
7/GUU/V	8/GCU/A	17/GAU/D	8/GGU/G
5/GUC/V	10/GCC/A	8/GAC/D	15/GGC/G
2/GUA/V	7/GCA/A	24/GAA/E	1/GGA/G
15/GUG/V	24/GCG/A	10/GAG/E	8/GGG/G

Tabla IV: Uso de codones del gene glnG de E. coli. Los aminoácidos se designan empleando el código de una letra.

niae fue analizado y comparado con varias proteínas regulatorias procarióticas. Como resultado, se dividió al producto el gene glnG en 4 bloques: B, C, D y E. Tres de ellos (B, D y E) se consideran como dominios (31) (Figura 8).

a) Dominio B

Se extiende del aminoácido 1 al 118. Es homólogo al extremo -NH<sub>2</sub> terminal de varias proteínas regulatorias: ompR (32), dye (33), nusA (34) y phoB (52) de E. coli; spo0F (35) y spo0A (36) de B. subtilis; cheB y cheY (37) de S. typhimurium; dctd de R. meliloti (52) y virG de Agrobacterium tumefaciens (52).

Todas estas proteínas controlan sistemas que responden a estímulos ambientales como la limitación de nitrógeno o fosfato, el choque osmótico, la presencia de ácidos dicarboxílicos de 4 C, compañeros sexuales, atrayentes y compuestos tóxicos. Todas interactúan con otras proteínas de control y regulan algunos genes a nivel de transcripción (38,40,41,42).

b) Bloque C

El bloque C, que es homólogo entre nifA de K. pneumoniae y de R. meliloti, glnG de E. coli y K. pneumoniae, ompR y dye de E. coli y spo0A de B. subtilis, tiene las propiedades de un segmento expuesto entre dominios. Es muy corto (19 aminoácidos), muy hidrofílico, su estructura secundaria predicha es de espiral y está localizado en el extremo -COOH del dominio B. La mayoría de sus residuos son glutamina, serina, ácido glutámico, arginina y prolina.

		C	D	E
Rm <u>nifA</u>		175 193	199	439 479 537
Kp <u>nifA</u>		135 205	211	450 451 520
Ec <u>glnG</u>	B	1 118	120 138	139 380 405 468
Kp <u>glnG</u>		1 118	120 138	139 380 405 469
Ec <u>ompR</u>		1 118	120 139	
Ec <u>dye</u>		1 116	118 137	
Bs <u>spo0A</u>		1 121	123 142	
Bs <u>spo0F</u>		1 118		
St <u>cheY</u>		1 102		
St <u>cheB</u>		1 100		
Ec <u>nusA</u>		325	371	

FIGURA 8: Comparación de los dominios de glnG de E. coli con los de otras proteínas regulatorias procarióticas. Las proteínas se alinearon por su homología en la secuencia de aminoácidos. Los números definen el primero y el último residuo de cada bloque. La barra listada indica un posible motivo que se une a DNA (30). Ec= E. coli; Kp= K. pneumoniae; Rm= R. meliloti; Bs= B. subtilis; St= S. typhimurium.



c) Dominio D

El dominio D es altamente homólogo (54%) entre los genes nifA de K. pneumoniae y R. meliloti y los genes glnG de E. coli y K. pneumoniae. Debido a esto, es posible que este dominio intervenga en la función de estos genes como activadores de la transcripción, interactuando con el producto del gene glnF o con la RNA polimerasa.

d) Dominio E

El dominio E se extiende del residuo 405 al 468 del producto del gene glnG. Es homólogo al extremo -COOH terminal de los genes nifA. Este dominio presenta el motivo característico de las proteínas que se unen a DNA.

## D I S C U S I O N

El resultado de este trabajo es la obtención de la secuencia nucleotídica de los genes glnL y glnG de E. coli, cuyos productos juegan un papel muy importante en la regulación del metabolismo del nitrógeno en esta bacteria.

La secuencia de aminoácidos deducida de la secuencia nucleotídica de estos genes, sugiere que son polipéptidos solubles, ya que no presentan dominios predominantemente hidrofóbicos.

En cuanto al uso de codones "no óptimos" de los genes glnL y glnG, presentan una incidencia característica de las proteínas regulatorias (mayor al 20%) (45).

La secuencia nucleotídica de estos genes y la deducida de aminoácidos está altamente conservada entre E. coli y K. pneumoniae (28,29,30), lo cual indica las relaciones evolutivas entre ambos organismos.

Por trabajos genéticos previos, se involucró al extremo -NH<sub>2</sub> terminal del producto del gene glnL en la regulación negativa de la expresión de glnA. Esto se logró mapeando mutaciones supresoras de cepas glnF<sup>-</sup> y mutaciones resultantes en una derepresión elevada de GS (15,21). Al abalizar la secuencia de aminoácidos deducida de la secuencia nucleotídica de esta región, se encontró que existe un dominio homólogo al que presentan las proteínas que se unen al DNA de manera específica. Este dato sugiere que la regulación negativa de la expresión de glnA se produce por la unión del producto de glnL al DNA. Esta hipótesis se debe probar experimentalmente.

En cuanto al producto del gene glnG, mutaciones en el

dominio B reducen la expresión de GS (15). Este dominio es homólogo al extremo  $-NH_2$  terminal de varias proteínas regulatorias (30), algunas implicadas en control positivo transcripcional. Por esto se pensó que este dominio estaba involucrado en regulación transcripcional. Pero algunas proteínas (como son los productos de los genes cheB y cheY (37)) no poseen tal función. Además, en el caso del producto de cheY, el cual sólo comprende este dominio (37), éste está implicado en la unión de S-adenosil metionina. Por lo tanto, no es probable que este dominio interactúe directamente con el aparato transcripcional, sino con nucleótidos de adenina o con otras proteínas.

En cuanto al dominio D, está involucrado en la función de activador de los productos de nifA y glnG. Mutaciones en esta región del gene de glnG eliminan la función positiva del producto de glnG, más no la negativa (15). Es probable que este dominio interaccione con el producto del gene glnF o con la RNA polimerasa.

Se ha reportado que el producto del gene glnG se une a DNA como represor (16,17,18). Por estudios genéticos, se ha visto que mutaciones que mapean en el dominio E eliminan la función de represor. En este dominio se encuentra una región homóloga a la implicada en la unión de proteínas a DNA, por lo que se sugiere que su función como represor se lleva a cabo por su unión a DNA.

Tanto las proteínas homólogas al producto de glnL como las que lo son al producto de glnG son proteínas regulatorias que controlan sistemas que responden a estímulos ambientales.

La mayoría de los miembros homólogos al extremo  $-NH_2$  terminal del producto de glnG se piensa que tienen funciones de

activadores transcripcionales, que actúan en concierto con los productos génicos homólogos al -COOH terminal del producto de glnL.

Se puede plantear un modelo de la interacción entre estos dos grupos de proteínas (52):

El -NH<sub>2</sub> terminal de las proteínas del grupo homólogo al producto de glnL percibirían señales ambientales específicas, las que transmitirían a la región conservada (-COOH terminal) a través de una modificación alostérica. Esta región conservada interacciona con o modifica al -NH<sub>2</sub> terminal de las proteínas homólogas al producto de glnG; la modificación afectaría la función del -COOH terminal (en el caso de glnG su función de activador o represor).

Se necesita más trabajo para poder asignar funciones exactas a los dominios que se bosquejan en esta investigación. El análisis presentado servirá para diseñar estos experimentos. La homología entre estos dos grupos de proteínas y su interacción, pudieran reflejar la integración de los sistemas de control a otro nivel.

#### Diferente expresión de los productos de los genes glnL y glnG de E. coli

En experimentos con minicélulas transformadas con plásmidos conteniendo todo el operón glnALG, nunca se ha detectado el producto de glnL (44). Además, en el caso de pACR8 en el que sí se observa, éste siempre se encuentra en menor cantidad que el de glnG, a pesar de que ambos se están transcribiendo del promotor antitet del vehículo donde está clonado.

Existen básicamente 3 mecanismos para explicar este fenómeno:

1) que se produzca menos producto de glnL porque el inicio de la traducción del producto de glnL sea menos eficiente que el de glnG debido a que su sitio de unión a ribosomas esté oculto o sea menos homólogo al consenso; o bien, que el paso de elongación de la cadena polipeptídica sea limitante, por que el gene presente un uso muy elevado de codones "no óptimos"

2) que el RNA mensajero de glnL se degrade

3) que el producto de glnL sea más inestable que el de glnG.

Para explorar la primera posibilidad, se practicó un análisis de incidencia de 23 codones "no óptimos" en los genes glnL y glnG, encontrándose que ambos genes presentan un uso elevado (mayor al 20%), igual a lo reportado para proteínas regulatorias (45). Por lo tanto, no parece plausible que esta sea la explicación de la diferente expresión de los productos de glnL y glnG. Quedan por descartar las otras dos hipótesis.

## A P E N D I C E

Sanger et al. (46,47) han desarrollado un método para secuenciar DNA el cual se basa en la propiedad del fragmento Klenow de la DNA polimerasa I de sintetizar una copia radioactiva complementaria de un templado de DNA de cadena sencilla usando un pequeño fragmento de DNA como "primer". El procedimiento implica el uso de los análogos 2'-3' dideoxi de los dideoxinucleósidos trifosfatados normales como terminadores de la polimerización. La síntesis de DNA se lleva a cabo en presencia de los cuatro deoxinucleótidos trifosfatados, uno de los cuales está marcado radioactivamente, o bien, se marca el "primer" en el extremo 5'. En cada reacción se añade uno de los cuatro dideoxinucleótidos, por lo que se requiere hacer cuatro reacciones. Por lo tanto, en cada reacción hay una población de moléculas de DNA radioactivas parcialmente sintetizadas, teniendo un extremo 5' común, pero variando en longitud hacia el extremo 3', dependiendo del sitio donde se haya insertado el dideoxinucleótido. Las cadenas sintetizadas se separan por tamaño por electroforesis en un gel de poliacrilamida-urea. Las cuatro reacciones se aplican al gel en carriles contiguos. Se autorradiografía el gel y se puede leer la secuencia obtenida.

El método requiere de templados de DNA de cadena sencilla. Estos son preparados al clonar los fragmentos a secuenciar en el genoma del bacteriófago M13 (48). Este fago se ha diseñado genéticamente para contener parte del gene de la enzima beta-galactosidasa de E. coli y expresa este gene al crecer en una cepa lac<sup>-</sup> (JM 101 ó 103). La región lac sirve como un sistema

para distinguir el DNA del vector (que dá origen a placas azules en cajas conteniendo 5-bromo-4-cloro-3-indolil-beta-D-galactopiranosido o X-gal) y el DNA recombinante (que forma placas blancas.

En esta región lac se ha introducido un inserto conteniendo 13 sitios únicos de restricción (49). Así, el fragmento a secuenciar se inserta en el vector M13 usando estos sitios y la clona resultante es fácilmente distinguible como una placa blanca.

Para obtener el templado adecuado, se toma una placa blanca y se crece en un pequeño volumen de medio de cultivo (YT). Se centrifuga para remover las bacterias. Durante su ciclo de replicación, el genoma del fago M13 presenta una forma replicativa de DNA de doble cadena y una forma de cadena sencilla, la cual es exportada de la bacteria envuelta en proteínas. De estas partículas es posible extraer el DNA de cadena sencilla. Usando un oligonucleótido sintético como "primer" universal (49), el cual hibridiza justo antes de la región de inserción, se puede obtener la secuencia del fragmento por el método de Sanger arriba descrito.

Los fragmentos de DNA a secuenciar se purificaron de geles de agarosa de bajo punto de fusión (1% en TBE 1X). Este DNA se clonó en los plásmidos M13 mp18 y mp19 digeridos previamente con la enzima de restricción apropiada. Se transfectó la cepa JM 101 con el DNA producto de la reacción de ligasa. Esta cepa se crece por dos horas en medio YT a 37°C con agitación, hasta que alcance una densidad óptica de 0.4 a 0.6 leída a 600 nm. Las células se centrifugan a 6 krpmX5'. Se resuspenden en 10 CaCl<sub>2</sub> 50 mM y se dejan en hielo por 30'. Se centrifugan de nuevo y se resuspende en 5 ml de la solución anterior. Se guardan a 4°C.

Se ponen 0.2 ml de células y 5 ul de la reacción de ligasa en tub

estéril. Se dejan en hielo por 45 min. Se tratan con calor a 42°C por 5 min. Se añaden 3 ml de agar suave, 25 ul de X-gal 2.5% en dimetilformamida y 25 ul de Isopropiltiogalactósido (IPTG) 2.5% en agua. Se plaquea en cajas de YT y se crece toda la noche a 37°C.

Se toma una de las placas blancas y se pone a crecer en 1.5 ml de una dilución 1/100 de un overnight de la cepa JM101. Se crece con agitación vigorosa (300 rpm) por 5 horas y media. Se transfiere el contenido del tubo a un tubo eppendorf y se centrifuga por 5 min a 10 krpm. El sobrenadante se transfiere a un tubo limpio, al que se agregan 150 ul de PEG (polietilen glicol 20% y cloruro de sodio 14.6% en agua) y se deja reposar a temperatura ambiente por 10 min. Se centrifuga por 5 min a 10 krpm. Se le quita todo el PEG. La pastilla se resuspende en 100 ul de TE (Tris 10 mM, EDTA 1mM pH 8.0). Se extrae con 50 ul de fenol saturado. Se añade medio volumen de acetato de amonio 7.5 M y 2.5 volúmenes de etanol al 95%. Se enfría a -70°C por 30 min. Se centrifuga a 10 krpm por 15 min. Se tira el sobrenadante y la pastilla se lava con un ml de etanol al 95%. Se centrifuga a 10 krpm por 4 min. y se tira el sobrenadante. La pastilla se seca y resuspende en 20 ul de TE. Se guardan a -20°C.

#### Secuencia:

Se marca el primer en el extremo 5' con  $\gamma$ -<sup>32</sup>P ATP

Primer universal de M13            0.4 pmol

$\gamma$ -<sup>32</sup>P ATP                            2.0 uCi

Polinucleótido kinasa            1.5 unidades

Buffer kinasa 10X                    1/10 de volumen de reacción final  
(0.7 M Tris pH7.6, 0.1 M MgCl<sub>2</sub>, 50 mM DTT)

Incubar a 37°C por 30 min.

Se hibridiza el primer con el templado:

5 ul de templado (0.5 a 1.0 ug)



1 ul de primer kinado (0.04 pmol)

1ul de buffer TM (Tris 100 mM, MgCl<sub>2</sub> 100 mM pH 8.5)

3 ul agua.

Se incuba a 60°C por una hora.

Se añaden 2 ul de la mezcla primer+templado en 4 tubos eppendorf a los que se adicionan 2 ul de la mezcla de deoxi-dideoxinucleótidos respectiva (G, A, T ó C) ver Tabla I\* y 2ul de la mezcla Klenow (Fragmento Klenow de la DNA Polimerasa I: 1 unidad, dATP 0.02 mM: 1 ul y 7 ul de agua).

Se deja a temperatura ambiente por 20 min. Se añaden 2 ul de "chase" que es una solución 0.25 mM de los cuatro deoxinucleótidos. Se incuba por 15 min. más. Se añaden 4 ul de Formamida desionizada (1 ml de formamida, 1 mg de xilen cianol ,y 1 mg de azul de bromo-fenol).

Se hierven los tubos por 3 min y se aplican a un gel de poliacrilamida

Tabla I\*

Mezclas de deoxi-dideoxinucleótidos

	T	C	G	A
0.5 mM dTTP	25	500	500	500
0.5 mM dCTP	500	25	500	500
0.5 mM dGTP	500	500	25	500
10 mM ddTTP	50			
10 mM ddCTP		8		
10 mM ddGTP			16	
10 mM ddATP				3
TE	1000	1000	1000	500

## GEL DE POLIACRILAMIDA - UREA

- Solución A: 7 ml de Solución I  
9.3 ul de Persulfato de Amonio 25%  
14 ul de TEMED
- Solución B: 35 ml de Solución II  
59 ul de Persulfato de Amonio 25%  
90 ul de TEMED
- Solución I: 150 ml de acrilamida desionizada al 40%  
50 ml de TBE 10X  
460 gr de urea  
50 mg de azul de bromofenol  
Aforar a un litro con agua.
- Solución II: 150 ml de acrilamida desionizada al 40%  
50 ml de TBE 10X  
460 gr de urea  
Aforar a un litro con agua.
- TBE 10X: 108 gr de Tris base  
55 gr de ácido bórico  
9.3 gr de EDTA  
Aforar a un litro con agua.

Se adapta una propipeta a una pipeta de 10 ml.

Se toman 4 ml de Solución B y luego 6 ml de Solución A.

Se introducen 5 burbujas de aire para formar un gradiente.

Se vacía a los vidrios de electroforesis y se añade el resto de la Solución B.

Se coloca el peine y se deja polimerizar por más de 30 min.

Se monta el gel en la cámara de electroforesis usando buffer TBE 1X para la corrida.

Se cargan 3 ul de cada reacción con un capilar con la punta angostada con calor.

Se corre a wattage constante (40 Watts) por 2 horas.

Se fija en ácido acético 10% y metanol 10% por 15 min.

El gel se transfiere a papel Whatman 3MM y se cubre con película plástica y se seca en un secador de geles a 80°C.

Se remueve la película plástica y se pone en un cassette de autorradiografía con una hoja de film para rayos X en contacto directo con el gel.

Se expone toda la noche y se revela al día siguiente.

## REFERENCIAS

- 1) Kustu, S., Burton, D., García, E., McCarter, L. & McFarland, N.  
1979. Proc. Natl. Acad. Sci. USA 76:4576
- 2) Magasanik, B.  
1982. Ann. Rev. Genet. 16:135-168
- 3) Merrick, M.  
1982. Nature 297:362-363
- 4) McFarland, N., McCarter, L., Artz, S. & Kustu, S.  
1981. Proc. Natl. Acad. Sci. USA 78:2135-2139
- 5) Rocha, M., Vázquez, M., Garcíarrubio, A. & Covarrubias, A.  
1985. Gene 37:91-99
- 6) Krajewska-Grynkiewicz, K. & Kustu, S.  
1983. Mol. Gen. Genet. 192:189-197
- 7) Pahel, G., Rothstein, D.M. & Magasanik, B.  
1982. J. Bact. 150:202-213
- 8) García, F., Bancroft, S., Rhee, S.G. & Kustu, S.  
1977. Proc. Natl. Acad. Sci. USA 74:1662
- 9) Gaillardin, C.M. & Magasanik, B.  
1978. J. Bact. 133:1329
- 10) Leonardo, J.M. & Goldberg, R.B.  
1980. J. Bact. 142:99
- 11) Hirschman, J., Wong, P.-K., Sei, K., Keener, J. & Kustu, S.  
1985. Proc. Natl. Acad. Sci. USA 82:7525-7529
- 12) Hunt, T.P. & Magasanik, B.  
1985. Proc. Natl. Acad. Sci. USA 82:8453-8457
- 13) Kustu, S., McFarland, N.C., Hui, S.P., Esmon, B. & Ames Ferro-Luzzi, G.  
1979. J. Bact. 138:218-234

- 14) Ueno-Nishio, S., Backman, K.C. & Magasanik, B.  
1983. J. Bact. 153:1247-1251
- 15) MacNeil, T., Roberts, G.P., MacNeil, D. & Tyler, B.  
1982. Mol. Gen. Genet. 188:325-333
- 16) Ames Ferro-Luzzi, G. & Nikaïdo, K.  
1985. EMBO J. 4:539-547
- 17) MacNeil, T., MacNeil, D. & Tyler, B.  
1982. J. Bact. 150:1302-1313
- 18) Reitzer, L.J. & Magasanik, B.  
1983. Proc, Natl. Acad. Sci. USA 80:5554-5558
- 19) Dixon, R.  
1984. Nucl. Acid, Res. 12:7811-7830
- 20) Ueno-Nishio, S., Mango, S., Reitzer, L. J. & Magasanik, B.  
1984. J. Bact. 160:379-384
- 21) Chen, Y. -M., Backman, K. & Magasanik, B.  
1982. J. Bact. 150:214-220
- 22) Reitzer, L. J. & Magasanik, B.  
1985. Proc. Natl. Acad. Sci. USA 82:1979-1983
- 23) Backman, K. C., Chen, Y. -M., Ueno-Nishio, S. & Magasanik, B.  
1983.. J. Bact. 154:516-519
- 24) Wei, G. R. & Kustu, S.  
1981. Mol. Gen. Genet. 183:392-399
- 25) Maniatis, T., Fritsch, E. F. & Sambrook, J.  
Molecular cloning: a laboratory manual. Cold Spring Harbor  
Laboratory. Cold Spring Harbor, N. Y. 1982.
- 26) Sanger, F., Nicklen, S. & Coulson, A. R.  
1982. Proc. Natl. Acad. Sci. USA 74:5463-5467
- 27) Covarrubias, A. A. & Bastarrachea, F.  
1983. Mol. Gen. Genet. 190:171-175

- 28) MacFarlane, S. A. & Merrick, M.  
1985. Nucl. Acid. Res. 13(21):7591-7606
- 29) Buikema, W. J., Szeto, W.W., Lemley, P. V., Orme-Johnson, W. H. & Ausubel, F. M.  
1985. Nucl. Acid. Res. 13(12):4539-4555
- 30) Drummond, M., Whitty, P. & Wootton, J.  
1986. EMBO J. 5(2):441-447
- 31) Rossmann, M. G. & Argos, P.  
1981. Ann. Rev. Biochem. 50:497-532
- 32) Wurtzel, E. T., Chou, M. -Y. & Inouye, M.  
1982. J. Biol. Chem. 257:13691-13695
- 33) Drury, L. S. & Buxton, R. S.  
1985. J. Biol. Chem. 260:4236-4242
- 34) Ishii, S., Ihara, M., Maekawa, T., Nakamura, Y., Ughida, H. & Imamoto, F.  
1984. Nucl. Acid. Res. 12:3333-3342
- 35) Trach, K., Chapman, J. W., Piggot, P. & Hoch, J. A.  
1985. Proc. Natl. Acad. Sci. USA 82:7260-7264
- 36) Ferrari, F. A., Trach, K., Le Coq, D., Spence, J., Ferrari, E. & Hoch, J. A.  
1985. Proc. Natl. Acad. Sci. USA 82:2647-2651
- 37) Stock, A., Koshland, D. E. Jr. & Stock, J.  
1985. Proc. Natl. Acad. Sci. USA 82:7989-7993
- 38) Hall, M. N. & Silhavy, T. J.  
1981. J. Mol. Biol. 151:1-15
- 39) Piggot, P. J.  
1985. En Dubnau, D. A. (ed.) The molecular biology of the bacilli. Academic Press.
- 40) Gaffney, D., Skurray, R. & Willets, N.

1983. J. Mol. Biol. 168:103-122
- 41) Ollington, J. F., Haldenwang, W. G., Huynh, T. U. & Losick, R.  
1981. J. Bact. 147:432-442
- 42) Hoch, J. A., Trach, K., Kawamura, F. & Saito, H.  
1985. J. Bact. 161:552-555
- 43) Pabo, C. O. & Sauer, R.T.  
1984. Ann. Rev. Biochem. 53:293-321
- 44) Rocha, M.  
Tesis doctoral.
- 45) Konigsberg, W. & Nigel Godson, G.  
1983. Proc. Natl. Acad. Sci. USA 80:687-691
- 46) Sanger, F., Nicklen, S. & Coulson, A. R.  
1977. Proc. Natl. Acad. Sci. USA 74:5463
- 47) Sanger, F.  
1981. Science 214:1205
- 48) Messing, J., Gronenborn, B., Muller-Hill, B. & Hofschneider, P. H.  
1977. Proc. Natl. Acad. Sci. USA 74:3642
- 49) Messing, J., Crea, R. & Seeburg, P. H.  
1981. Nucl. Acid. Res. 9:309
- 50) Ohlendorf, D. H. & Matthews, B. W.  
1983. Ann. Rev. Biophys. Bioeng. 12:259-284
- 51) Ikemura, T. & Ozeki, H.  
1983. C.S.H.S.Q.B. 47:1087-1097
- 52) Tracy Nixon, B., Ronson, C. W. & Ausubel, F. M.  
1986. En Abstracts of the Third International Symposium on the molecular genetics of plant-microbe interactions. p. 207
- 53) Dayhoff, M. O., Hunt, L. T. & Hurst-Calderone, S.

1978. En Atlas of Protein Sequence and structure.

M. O. Dayhoff ed. Vol. 5 suppl. 3 pp. 363-369