

03067
2eje 6

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Unidad Académica de los Ciclos Profesional y de Postgrado
del CCH

Centro de Investigación sobre Ingeniería Genética
y Biotecnología
Cuernavaca, Morelos

Caracterización de los genes que codifican para la enzima
glutamato sintasa de Escherichia coli K-12

T E S I S

Que para obtener el título de:

Maestro en Investigación Biomédica Básica

Presenta

Guillermo Gosset Lagarda

1988

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

INDICE

NOTA ACLARATORIA.....	1
ANTECEDENTES Y OBJETIVOS.....	2
INTRODUCCION.....	7
MATERIALES Y METODOS.....	13
RESULTADOS.....	14
"Determination of the nucleotide sequence for the glutamate synthase structural genes of <u>Escherichia coli</u> K-12" Oliver G., Gosset G., Sanchez-Pescador R., Lozoya E., Ku L. M., Flores N. Becerril B., Valle F. y Bolivar F.....	15
"Amino acid sequence analysis of the glutamate synthase enzyme from <u>Escherichia coli</u> K-12" Gosset G., Merino E., Recillas F., Oliver G., Becerril B. y Bolivar F.....	26
RESULTADOS ADICIONALES Y DISCUSION.....	59
PERSPECTIVAS.....	81
BIBLIOGRAFIA ADICIONAL.....	85

NOTA ACLARATORIA:

La mayor parte de los resultados obtenidos durante el desarrollo de esta tesis fueron integrados en dos artículos. Es necesario resaltar la participación de las personas que aparecieron como coautores en los mismos, sus nombres aparecen en las publicaciones respectivas. Dichas publicaciones se anexan en la sección de RESULTADOS. En la sección RESULTADOS ADICIONALES Y DISCUSION se incluyen datos no comprendidos en los dos artículos y se discute la información total del trabajo de tesis.

ANTECEDENTES Y OBJETIVOS

Uno de los principales intereses en nuestro laboratorio es el estudio y la caracterización de regiones específicas del genoma bacteriano, así como de los productos codificados por dichas regiones. Por este motivo, se han aislado los genes que codifican para enzimas que intervienen en el metabolismo nitrogenado de enterobacterias (Covarrubias et al., 1980), así como el gene de la enzima penicilino acilasa de *E. coli* (Oliver et al., 1986). Estos genes han sido caracterizados a un nivel muy fino, lográndose tener una idea clara de su organización, así como de los elementos que pudieran participar en la regulación de su expresión (Oliver et al., 1987; Valle et al., 1986; Riba et al., 1988).

Las técnicas utilizadas para estos estudios, descansan en conceptos teóricos de la Biología Molecular y la Fisicoquímica de Macromoléculas. Estas técnicas han evolucionado rápidamente, lográndose avances en cuanto a la eficiencia y la precisión con la que se puede obtener información de sistemas biológicos. El tipo de información que posiblemente se genera con mayor rapidez es el de las secuencias nucleotídicas y de aminoácidos. Esta información es acumulada en bancos de datos, facilitándose de esta manera que puedan ser comparadas y relacionadas entre sí

utilizando programas para computadora. Para tener una idea de la relevancia de la información contenida en estos bancos, es importante considerar la propuesta de un origen común para la mayoría de los genes existentes y por lo tanto de las moléculas codificadas por los mismos. En el caso de las secuencias de aminoácidos, es posible obtener información de las mismas, comparándolas con secuencias de proteínas cuya estructura tridimensional y sitios de unión a diversas moléculas ya ha sido determinado experimentalmente. Estas comparaciones permiten la predicción de propiedades o funciones en un segmento proteico. Actualmente la mayor parte de las secuencias de aminoácidos que se reportan, provienen de la traducción de secuencias nucleotídicas de genes estructurales. Por esta razón, es frecuente que se desconozca cuales regiones de las proteínas participan en la catálisis o regulación alostérica, cuando estas existen, e incluso algunas veces se ignora cual puede ser la función del polipéptido codificado por la respectiva secuencia nucleotídica.

Una de las líneas de investigación en nuestro laboratorio es el estudio de los genes que codifican para dos de las enzimas involucradas en la asimilación de amonio en *E. coli*: la glutamato deshidrogenasa (GDH) y la glutamato sintasa (GOGAT). La asimilación de amonio bajo distintas condiciones metabólicas resulta de la acción de estas dos enzimas y la enzima glutamato sintetasa (GS), además de otras enzimas que participan en el metabolismo de compuestos nitrogenados. La enzima GOGAT cataliza la síntesis de glutamato utilizando los sustratos

alfa-cetoglutarato y glutamina, siendo esta reacción un punto de cruce para los metabolismos del nitrógeno y del carbono. Debido a la importancia de esta enzima dentro del metabolismo celular, se ha desarrollado una estrategia de estudio cuyos resultados más relevantes hasta antes de esta tesis se mencionan a continuación:

- Aislamiento del plásmido PRSP20 el cual contiene un fragmento de DNA cromosomal capaz de complementar la mutación *gltB31* de E. coli.
- Identificación de dos polipéptidos codificados por el plásmido PRSP20 que son reconocidos por anticuerpos específicos para las dos subunidades de GOGAT.
- Determinación de la dirección de transcripción para los dos genes estructurales de GOGAT.
- Determinación de la secuencia nucleotídica del gene estructural para la subunidad grande y regiones flanqueadoras, secuencia nucleotídica parcial del gene estructural para la subunidad pequeña.
- Determinación de la secuencia de los primeros veinte aminoácidos de la región amino terminal para las dos subunidades.

- Identificación por criterio de consenso de secuencias en los genes estructurales para GOGAT y regiones flanqueadoras, con un posible papel regulatorio.

Los tres primeros puntos ya han sido reportados en la literatura por nuestro laboratorio (Covarrubias et al., 1980; Lozoya et al., 1980; Garcarrubio et al., 1983). De las tres enzimas que participan en la asimilación de amonio, GOGAT es la que se encuentra menos estudiada en lo que respecta a su organización genética a nivel molecular. Este trabajo pretende aportar información sobre la organización transcripcional de los genes que codifican para GOGAT, y además establecer relaciones entre lo que se conoce sobre la estructura y mecanismos de acción de la enzima y su secuencia de aminoácidos. Esta tesis es una continuación de la caracterización de los genes estructurales para GOGAT y sus productos. Sus objetivos fueron los siguientes:

OBJETIVOS DE LA TESIS

- Completar la secuencia nucleotídica del gene estructural para la subunidad pequeña de GOGAT (gltD).
- Identificar la(s) región(es) de inicio de la transcripción para los dos genes estructurales de GOGAT (gltB y gltD).
- Identificar el (los) transcrito(s) para gltB y gltD.

- Localizar secuencias posiblemente involucradas en la regulación de los genes que codifican para GOGAT.

- Realizar un análisis comparativo de la secuencia de aminoácidos deducida para las dos subunidades de GOGAT, basado en la identificación de similitudes con secuencias de proteínas compiladas en un banco de datos.

INTRODUCCION

La asimilación de amonio en E. coli depende de la acción coordinada de las enzimas glutamino sintetasa (GS), glutamato deshidrogenasa (GDH) y glutamato sintasa (GOGAT). Estas enzimas catalizan la síntesis de los aminoácidos glutamina (GS) y glutamato (GDH y GOGAT), los cuales aparte de ser incorporados directamente a las proteínas, han sido seleccionados como las moléculas donadoras de grupos amido y amino para la síntesis de compuestos nitrogenados en la célula (Tyler, 1978).

El crecimiento de E. coli en un medio mínimo con una baja concentración de amonio (<0.5 mM), requiere de la asimilación de este compuesto por la acción de las enzimas GS y GOGAT. En estas condiciones, GS incorpora el amonio al glutamato, para formar glutamina. Este aminoácido es a su vez utilizado por GOGAT como uno de sus substratos, transfiriendo el grupo amido de la glutamina al 2-oxoglutarato, produciendo así dos moléculas de glutamato. Una parte del glutamato sintetizado por GOGAT puede reciclarse y volver a ser utilizado como substrato por la GS. De esta forma, el amonio es repartido entre la glutamina y el glutamato en un ciclo que permite un rápido recambio entre estos dos compuestos.

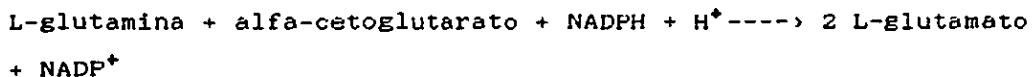
Desde su descubrimiento en K. aerógenes por Tempest en 1970, la enzima GOGAT ha sido identificada en diversos grupos biológicos incluyendo bacterias (Tempest et al., 1970), algas

(Lea y Mifflin, 1975), hongos (Roon et al., 1974) y plantas superiores (Mifflin y Lea, 1976). Estos organismos tienen en común, la capacidad de crecer utilizando algún tipo de nitrógeno inorgánico, lo cual puede ser una indicación de que GOGAT contribuye a la asimilación de ese nutriente cuando se encuentra en concentraciones bajas en el medio. Mutantes deficientes en GOGAT, en E. coli y otras enterobacterias estudiadas, pueden crecer normalmente en un medio mínimo que contenga glucosa y un exceso de amonio. Sin embargo, son incapaces de crecer si existe una limitación de nitrógeno, a menos que se suministre glutamato al medio (Tyler).

Debido a su importante papel, dado que cataliza una reacción que es punto de cruce entre el metabolismo nitrogenado y el del carbono, la enzima GOGAT y las regiones de DNA que la codifican, deben poseer mecanismos de control finos y complejos, que pudieran responder a los niveles intracelulares de nitrógeno y carbono (Rosenfeld et al., 1973). Estudios en enterobacterias han revelado que la actividad de GOGAT puede variar bajo ciertas condiciones de crecimiento. Se ha detectado una disminución de la actividad de GOGAT, al crecer al organismo en un medio que contenga glutamato o algún aminoácido que pueda dar origen a ese compuesto. En B. subtilis y en E. coli se ha demostrado que la menor actividad de GOGAT en condiciones de crecimiento con glutamato en el medio de cultivo, corresponde a una disminución en el nivel del transcrito específico para las dos subunidades (Bohannon, 1985; Castaño et al., 1988). En E. coli, Prusiner et al., (1972) ha reportado una disminución en el nivel de la

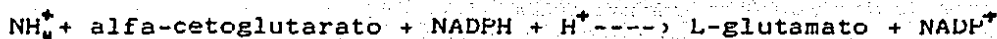
enzima, al aumentar los niveles intracelulares de AMPc. Por otro lado, Mora et al. (1987) han demostrado una inhibición alostérica por intermediarios del ciclo de los ácidos tricarboxílicos en la enzima GOGAT de *N. crassa*. Miller y Stadtman (1972) reportaron la inhibición *in vitro* de la actividad de la GOGAT de *E. coli* por los aminoácidos D y L-aspartato, D-glutamato y L-metionina. También demuestran una inhibición por oxaloacetato, un intermediario en el ciclo de los ácidos tricarboxílicos, aunque se debe mencionar que tal inhibición aparentemente es irreversible. Los datos anteriores indican que GOGAT puede ser regulada tanto en su actividad enzimática como en el nivel de síntesis de su RNA mensajero. Siendo los niveles de nitrógeno y carbono los que pudieran modular su actividad y su expresión.

La enzima GOGAT presenta una serie de características que la hacen interesante por su papel dentro del metabolismo nitrogenado, su actividad enzimática y su complejidad estructural. Por esta razón, se ha estudiado en detalle el mecanismo de acción de esta enzima. La reacción dependiente de glutamina catalizada por GOGAT es la siguiente:



(Reacción A)

Además, se ha demostrado que GOGAT también cataliza la síntesis de glutamato dependiente de amonio de acuerdo a la siguiente reacción:



(Reacción B)

Esta actividad solo representa el 5-7% de la actividad dependiente de glutamina. La reacción B es muy parecida a la reacción catalizada por la glutamato deshidrogenasa de E. coli dependiente de NADPH, siendo sin embargo, las K_m para los sustratos, distintas en las dos enzimas. Es importante mencionar que existen datos que indican que la reacción B puede ser catalizada exclusivamente por la subunidad pequeña de GOGAT (Mantsala and Zalkin, 1976b).

En E. coli K-12 esta enzima está compuesta de dos subunidades con un peso molecular de aproximadamente 116 kDa (1514 aa) y 52 kDa (417 aa) para las subunidades grande y pequeña respectivamente (Oliver et al., 1987). GOGAT es una flavoproteína que posee dos grupos hierro-azufre, estudios de velocidad de sedimentación sugieren que las flavinas (FAD y FMN) y los dos

grupos fierro-azufre se encuentran en la subunidad grande de la enzima (Trotta et al., 1974). Además, análisis realizados con la apoenzima, indican que la síntesis de glutamato dependiente de glutamina, requiere la presencia de los grupos prostéticos antes mencionados (Mantsala y Zalkin, 1976b).

Utilizando análogos de glutamina se ha demostrado que estos compuestos se unen a un residuo de cisteina en la subunidad grande de GOGAT, ocasionando la inhibición de la actividad dependiente de glutamina (reacción A), pero no así la actividad dependiente de amonio (reacción B) (Trotta et al., 1974; Mantsala y Zalkin, 1976b).

Por el tipo de reacción que cataliza, GOGAT se clasifica dentro de las enzimas glutamino amidotransferasas. Estas enzimas generalmente están constituidas por dos subunidades, una de las cuales (la subunidad pequeña) posee un sitio de unión para glutamina y presenta actividad de glutaminasa. La subunidad grande posee la actividad de sintetasa, utilizando como uno de sus sustratos al grupo amido removido de la glutamina por la subunidad pequeña. Reportes previos han establecido semejanzas y diferencias entre GOGAT y otras glutamino amidotransferasas (Mantsala y Zalkin, 1976a; Trotta et al., 1974; Mantsala y Zalkin, 1976b; Geary y Meister, 1977). Al igual que estas enzimas, GOGAT posee una actividad de glutaminasa, además, puede utilizar amonio en lugar de glutamina para la síntesis de glutamato. Las diferencias más importantes entre GOGAT y el resto de las glutamino amidotransferasas son: la presencia de grupos

flavina y fierro-azufre, así como la unión de glutamina y la actividad de glutaminasa en la subunidad grande de GOGAT.

MATERIALES Y METODOS

Los materiales utilizados en este trabajo se describen en las publicaciones anexadas en la sección de RESULTADOS. Los métodos que se enlistan a continuación, también se describen en dichas publicaciones.

- Determinación de la secuencia nucleotídica de DNA
- Hibridación RNA-DNA (Northern)
- Extensión de primero
- Análisis de la secuencia nucleotídica
- Análisis de la secuencia de aminoácidos

RESULTADOS

GEN 02179

Determination of the nucleotide sequence for the glutamate synthase structural genes of *Escherichia coli* K-12

(Recombinant DNA; nitrogen metabolism; protein amino terminus; functional promoter; transcriptional start point; flavoprotein; oligodeoxynucleotide primer)

Guillermo Oliver*, Guillermo Gosset, Ray Sanchez-Pescador*, Edmundo Lozoya, Lallig M. Ku*, Noemi Flores, Baltazar Becerril, Fernando Valle and Francisco Bolivar

Departamento de Biología Molecular, Centro de Investigación sobre Ingeniería Genética y Biotecnología, U.N.A.M., Cuernavaca, Morelos (México)

Received 31 July 1987

Accepted 12 August 1987

SUMMARY

We have determined the complete nucleotide sequence of a 6.3-kb chromosomal *HpaI-EcoRI* fragment, that contains the structural genes for both the large and small subunits of the *Escherichia coli* K-12 glutamate synthase (GOGAT) enzyme, as well as the 5'- and 3'-flanking and intercistronic DNA regions. The M_s s of the two subunits, as deduced from the nucleotide (nt) sequence, were estimated as 166208 and 52246. Partial amino acid sequence of the GOGAT enzyme revealed that the large subunit starts with a cysteine residue that is probably generated by a proteolytic cleavage. Northern blotting experiments revealed a transcript of approximately 7300 nt, that at least contains the cistrons for both subunits. A transcriptional start point and a functional promoter were identified in the 5' DNA flanking region of the large subunit gene. The messenger RNA nontranslated leader region has 120 nt and shares identity with the leader regions of *E. coli* ribosomal operons, in particular around the so-called *boxA* sequence implicated in antitermination. Other possible regulatory sequences are described.

INTRODUCTION

The synthesis of glutamate and glutamine involves three primary enzymes: glutamate dehydrogenase,

glutamine synthetase, and glutamate synthase (GOGAT, glutamine amide 2-oxoglutarate amido-transferase, EC 2.6.1.53). Glutamate and glutamine not only feed directly into protein synthesis, but also

Correspondence to: F. Bolivar, Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, C.P. 62270 Cuernavaca, Morelos (México) Tel. (5273)17-23-99.

* Present addresses: (G.O.) Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90024 (U.S.A.), Tel. (213)206-1401; (R.S.-P. and L.M.K.) Chiron Research Laboratories, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608 (U.S.A.) Tel. (415)655-8729.

Abbreviations: aa, amino acid(s); bp, base pair(s); cAMP, cyclic AMP; CRP, cAMP-receptor protein; GOGAT, glutamate synthase; kb, kilobases or 1000 bp; NN, no nitrogen minimal medium; nt, nucleotide(s); ORF, open reading frame; PA, polyacrylamide; R, purine; RBS, ribosome-binding site(s); SD, Shine-Dalgarno sequence of RBS; SDS, sodium dodecyl sulfate; Y, pyrimidine; [], designates plasmid-carrier state.

serve as amino-group donors for a wide range of other nitrogen-containing cellular components. Because of their central position in metabolism, glutamate dehydrogenase and GOGAT enzymes and their structural genes may have unique control features in addition to those found for other biosynthetic pathways, as is the case for glutamine synthetase (Rosenfeld et al., 1983).

In enteric bacteria glutamate can be synthesized by a variety of reactions. Nevertheless, a key reaction among them is the one catalyzed by GOGAT (Tyler, 1978). Through this reaction, where an amido group from glutamine is transferred to 2-oxoglutarate, a Krebs cycle intermediate, nitrogen and carbon metabolism are brought together.

The M_r of bacterial, fungal and plant GOGAT enzymes has been reported to be approx. 800000 (Tyler, 1978; Hummelt and Mora, 1980). In *E. coli* it may exist as an aggregate of four catalytically active dimers. In *E. coli* W each dimer consists of two nonidentical subunits whose estimated M_r s are 135000 and 53000 (Miller and Stadman, 1972). In *Neurospora crassa* and in plants this enzyme is composed of four identical monomers with an estimated M_r of 220000, the sum of the two subunits in bacteria (Hummelt and Mora, 1980).

GOGAT is an iron-sulfur flavoprotein, in which the large subunit binds iron and sulfur (Mantsala and Zalkin, 1976a,b; Rendina and Orme-Johnson, 1978). Glutamine binds to the large subunit and transfers the amido group to 2-oxoglutarate that apparently binds to the small subunit (Mantsala and Zalkin, 1976c; Trotta et al., 1974). These two features distinguish GOGAT from other well-characterized bacterial glutamine amidotransferases, which do not possess iron-sulfur or flavin groups and which in some cases bind glutamine to the small subunit. This unique structure of GOGAT may reflect some hitherto unknown function in addition to its role in glutamate synthesis, since iron-sulfur structures are usually characteristic of enzymes involved in electron transport and oxidation-reduction reactions (Rosenfeld et al., 1983).

We have previously reported the cloning of the *E. coli* K-12 genes coding for the GOGAT enzyme using a ColE1 hybrid plasmid, pRSP20. We have also reported evidence which indicated that both genes are tightly linked (Covarrubias et al., 1980; Lozoya et al., 1980; Garcarrubio et al., 1983).

In this paper, we present the nucleotide sequence of the *E. coli* K-12 chromosomal DNA segment carried in pRSP20 that includes the two GOGAT structural genes and their flanking sequences. Sequencing the N-termini of the purified enzyme identified the first 20 aa residues for both subunits. A transcriptional start point was localized in the 5'-flanking region of the gene that codes for the large subunit. A functional promoter was also identified. Other possible regulatory signals are proposed.

MATERIALS AND METHODS

(a) Bacterial strains and plasmids

E. coli K-12 strain PA340 is a glutamate auxotroph (Berberich, 1972). This strain carries a deletion of the two genes that code for the GOGAT subunits (A. Covarrubias, personal communication). Strain CS520 was obtained from J. Carbon (Clarke and Carbon, 1975). Plasmid pRSP20 carries the genes that code for the two GOGAT subunits (Covarrubias et al., 1980).

(b) Enzymes, radiochemicals and oligodeoxynucleotides

Enzymes were purchased from New England Biolabs or Boehringer Mannheim and used as specified by the manufacturers. The labeled [γ - 32 P]ATP, [α - 32 P]ATP and [α - 32 P]CTP were obtained from Amersham. All synthetic oligodeoxynucleotides were synthesized and purified using the protocol described by Sanchez-Pescador and Urdea (1984).

(c) Nucleic acid sequencing

Several restriction fragments of plasmid pRSP20 were cloned into the vectors M13mp18 and M13mp19. Dideoxy chain-termination sequencing reactions were carried out according to the methods reported by Messing et al. (1981) and Sanchez-Pescador and Urdea (1984). The last method involves 'walking across' a small number of large fragments, using specific synthetic oligodeoxynucleotide primers. Over 95% of the sequence was determined twice or more.

(d) Protein purification and amino acid sequencing

GOGAT was purified to homogeneity as previously described by Sakamoto et al. (1975). Partial amino acid sequence determination was obtained using 120 pmol of the purified enzyme. This sample was loaded in an Applied Biosystem gas-phase sequencer (model 470a) and subjected to 20 cycles of Edman degradation. Two sequences were identified and could be distinguished for several cycles due to differences in the observed amount (110 and 90 pmol). These sequences were later identified as the N termini of the small and large subunits, respectively.

(e) Nucleotide and protein sequence analysis

Standard Pascal programs for an Apple II computer, described by De Banzie et al. (1984) and Fristensky et al. (1982) were used.

(f) Northern blotting procedures

E. coli K-12 strains PA340, PA340[pRSP20] and CS520, were grown in NN minimal medium supplemented with 15 mM or 0.5 mM NH_4Cl as the nitrogen source and 11 mM glucose as the carbon source, and respective strain requirements (Covarubias et al., 1980). Cells were harvested at late exponential phase. RNA was isolated by a hot-phenol method (Young and Furano, 1981), fractionated by 1% agarose-2.2 M formaldehyde gel electrophoresis, and transferred to a nitrocellulose membrane. DNA fragments of the pRSP20 plasmid were nick-translated and used as radioactive ^{32}P probes.

(g) Primer extension procedure

A modification reported by Leon et al. (1985) was followed in using a specific synthetic oligodeoxynucleotide for the large subunit gene as primer for the synthesis of cDNA.

RESULTS

(a) Nucleotide sequence of the GOGAT structural genes

The recombinant plasmid pRSP20, which carries the structural genes for both *E. coli* K-12 GOGAT subunits, has been previously described (Covarubias et al., 1980; Lozoya et al., 1980). This molecule was used as the source of DNA for the sequencing experiments reported here.

A restriction map of the 6.3-kb chromosomal DNA insert in pRSP20 is shown in Fig. 1 along with the strategy used to sequence it. Different M13mp derived clones carrying the complete GOGAT genes and flanking regions were obtained. Universal and specific synthetic primers were used in the nucleotide sequence reactions. These procedures enabled us to efficiently 'walk across' the entire DNA fragment coding for both GOGAT subunits.

Examination of all possible translation reading frames of the sequenced 6.3-kb *HpaI-EcoRI* DNA fragment (Figs. 1 and 2), identified only two major ORFs, each long enough to encode for the relevant GOGAT subunit. The first ORF (nt positions 236-4845 in Fig. 2), showed two methionine residues as good candidates for translation initiation of the large subunit (nt positions 301 and 394). Both residues are in the same ORF and have good putative RBS (Stormo et al., 1982).

The second ORF (nt positions 4846-6270 in Fig. 2), has a possible translation initiation site at the methionine residue located at nt position 4858. This assumption is also supported by the presence of a possible RBS sequence (nt position 4843).

(b) The N-terminal sequence and *M_r* determinations of the GOGAT subunits

To identify the mature N-terminal sequences of both GOGAT peptide subunits, a partial protein sequence of the purified enzyme was determined. Based on the amino acid sequence obtained for the first N-terminal 20 aa residues for the two subunits, we were able to determine that the mature large GOGAT peptide subunit begins with a cysteine residue localized at nt position 427 and the mature small subunit at the serine residue located at nt position 4861 (Fig. 2).

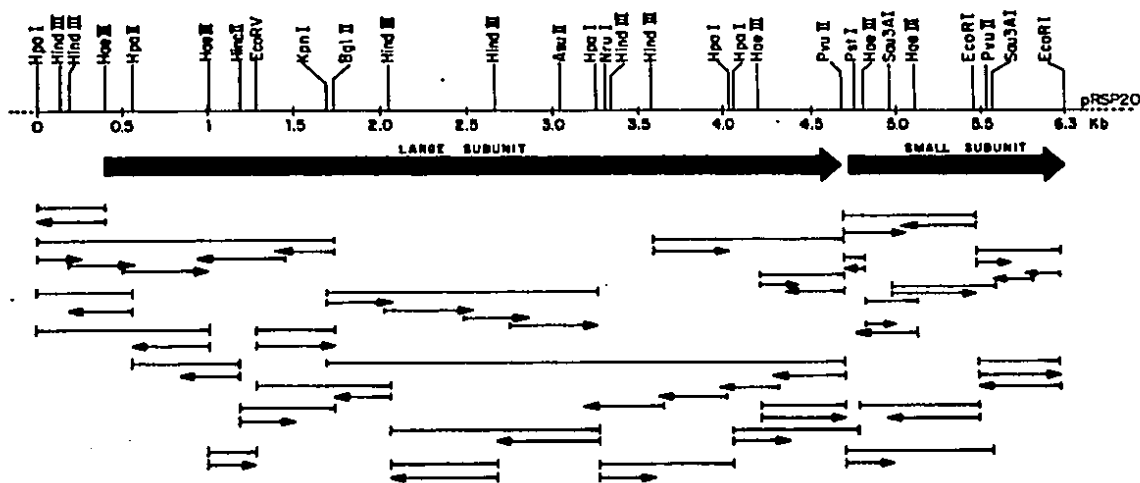


Fig. 1. Restriction map of *E. coli* GOGAT genes in plasmid pRSP20. Plasmid pRSP20 is a ColE1 derivative with an 8-kb *E. coli* CS520 chromosomal DNA insertion at the *EcoRI* site (Covarrubias et al., 1980; Lozoya et al., 1980; Garcarrubio et al., 1983). Only relevant restriction sites are indicated. Heavy arrows indicate the coding regions for the large and small subunits of the GOGAT enzyme. Below these heavy arrows, the sequencing strategy is shown. The M13 clones used are represented by lines above the thin arrows. These arrows indicate the direction and extent of nucleotide sequence determinations. The arrow tails indicate the location of the universal or specific synthetic primers used to sequence the corresponding stretch of DNA. The scale is in kb.

Based on these data, the first ORF encodes a protein of 1514 aa residues, with a predicted M_r of 166208, while the second ORF encodes a protein of 471 aa residues with a predicted M_r of 52246. Since it has been previously reported that the M_r of the large *E. coli* W GOGAT subunit is 135000 (Miller and Stadman, 1972), we decided to confirm the M_r obtained from the nucleotide sequence data. Fig. 3 shows an SDS-PA gel of purified *E. coli* K-12 GOGAT enzyme using *E. coli* β -galactosidase and RNA polymerase holoenzyme as M_r markers. The M_r for the large GOGAT subunit determined from this experiment (higher than 155000) is in close agreement with that deduced from the nucleotide sequence.

(c) Codon usage for both subunits

The analysis of the codon usage reveals the following features: The codons, CUA (for leucine), AUA (for isoleucine), and CGA/AGA/AGG (for arginine), corresponding to weakly interacting or minor tRNAs (Grosjean and Fiers, 1982), occurred

infrequently or not at all in both subunits. In accordance with the general trend for efficiently expressed genes in *E. coli*, the following codons were found to predominate: CUG for leucine, GGY for glycine, AUY for isoleucine, CCG for proline, GCG for alanine, GAA for glutamate, CGY for arginine, AAA for lysine, and AAC for asparagine.

(d) Northern blotting experiments and transcriptional start point determination

Two different nick-translation-labeled ^{32}P probes carrying structural regions of the corresponding genes for both GOGAT subunits were used to detect the corresponding mRNA. RNA was purified from *E. coli* strains CS520, PA340 and PA340 carrying plasmid pRSP20, grown in NN medium supplemented with glucose and NH_4Cl (see MATERIALS AND METHODS, section f), and transferred to a nitrocellulose filter. Hybridization using mRNA from strains CS520 and PA340[pRSP20] with either of the two probes, revealed only a single mRNA transcript of approx. 7.3-kb. When the same experiment

was done using RNA isolated from strain PA340, no band was obtained (Fig. 4). These results indicate that, at least under these growth conditions, a large polycistronic mRNA molecule is produced.

The transcription start point for this 7.3-kb mRNA molecule was determined by reverse transcriptase primer extension. Fig. 5 shows the result of such an experiment, where RNA isolated from PA340[pRSP20] cells grown on NN medium supplemented with glucose and 15 mM NH_4Cl , and a synthetic oligodeoxynucleotide (30 nt) that hybridizes at nt positions 473 to 503 (Fig. 2), were used. The size of the reverse transcription product was 325 ± 2 nt. This result located the transcription start point approximately at nt position 178 (Fig. 2). When the same experiment was done using RNA isolated from the control strain PA340, no band was observed (Fig. 5).

(e) Promoter localization and identification of potential transcriptional regulatory sequences

The nucleotide sequence preceding the DNA that codes for the N terminus of the large subunit was searched for similarities to sequences that are well conserved in *E. coli* promoters (Hawley and McClure, 1983). Two possible promoters were identified from consensus considerations. One of these putative promoters (nt position 146–174, Fig. 2) is localized immediately before the transcription start point, as determined by primer extension.

Other revealing features were observed in this nucleotide sequence preceding the DNA that codes for the N terminus of the large subunit. Two possible CRP-cAMP binding sites were detected following consensus criteria (deCrombrugge et al., 1984) (centered at nt positions 152–153 and 239–240, respectively, Fig. 2). A twofold symmetry region was also found beginning at nt position 197 (Fig. 2).

In addition, the nucleotide sequence starting at nt position 178 (Fig. 2) exhibits similarity to the *rmB* nucleotide sequence that codes for the ribosomal RNA leader sequence (Gourse et al., 1986). This region includes an antitermination sequence (*boxA*), similar to the consensus reported by Olson et al. (1984) (Fig. 6).

Finally, examination of the 3' DNA region that codes for the large subunit revealed a sequence with dyad symmetry (nt positions 4826–4841, Fig. 2).

DISCUSSION

The complete nucleotide sequence of the *E. coli* K-12 GOGAT-coding genes, including the inter-cistronic and the 5'- and 3'-flanking regions, has been determined. The first ORF (nt positions 236–4845, Fig. 2) revealed two methionine residues as good candidates for translation initiation. Both residues are preceded by putative RBS. Presently, we do not know which of these two sites is used. However, published evidence suggests that translation initiation could occur preferentially at the second methionine residue, due to the following facts: Gourse et al. (1986) have shown that A or U residues following the RBS are favorable for translation efficiency in *E. coli*, whereas G and to a lesser extent C residues are inhibitory to translation. Furthermore, an A residue at the -2 position enhances translation efficiency (De Boer et al., 1983). It has also been reported that the sequence ACA, as the second codon in the β -galactosidase mRNA, reduces translation efficiency, whereas the sequence UUG has an eight-fold increase in translation efficiency as compared with ACA (Hui et al., 1984). In accordance with these data, we propose that the second methionine residue (nt position 394, Fig. 2), is the most probable translation initiation site. Nevertheless, we do not rule out the possibility that the first methionine residue (nt position 301, Fig. 2), may work in certain conditions. Furthermore, several cases have been reported in which the translation of one mRNA could begin from two in-frame methionine residues (Normark et al., 1983).

Based on the N-terminal sequence of the two subunits of this enzyme, it was possible to determine that the mature large GOGAT subunit begins with a cysteine residue as indicated in Fig. 2. In accordance with these data, we propose that the mature N terminus of the large subunit is created by proteolytic cleavage of a precursor, whose translation start codon corresponds to one of the two suggested methionine residues. Usually, the proteins in *E. coli* which are posttranslationally processed have either a typical cleavable exporting leader peptide at the N terminus, or are created from a common polypeptide precursor (Oliver et al., 1985; Makaroff et al., 1983). However, it has been reported that the mature glutamine phosphoribosylpyrophosphate amidotransferase (EC 2.4.2.14) from *B. subtilis*, is created

GTTAAACACACTTATGACAGTCCAGAAATGACTGTTTCTCAACGACTTCCCTTTTAGCCCTAAAGATAAAATCCATTTTAAATTCAGCTATTAATAAAGAAATTTGCGCTAAAGCACA 120
 TTTCTGTACCAATAAGCTTGCATTT-35 -10 MW Box A
 TTTCTGTACCAATAAGCTTGCATTTAGCTGTATCAAGCTTCCCGATAAGTTGGAAATCCCGCTGGAAGCTTTCTGGATGACGACCTGCTCATATAATTTATGCGATAATTGAGATCCC 240
 CTCTTACCCTGTTAATACCGATGCGAAAGGCACAAAGGGGGCGAATCCGCGCCGCTATGACACGCAAAACCCCTGCGCACGCTCTTCTGCGCCGTCGCCAGCGGTTGGAAATG 360
 S.D. [2] T R R P R R A L S V P V R G S E V
 GGTTCCCGCAGACGCTCGCGGCTTCCAGCATATCTTGTACGATAAATCCCTCTGAGAGGGATAACTGCTGCTGCGCTGATGCCCCACATAGAAAGCCGAACTAGCCACAAGCTAGT 480
 G P F O S L G E V H D [2] L Y D R S L R L R G C F G L I A N I E G Z P S N K V V
 CGTACTAATACAGCCACTGCCCGCATGACGACCGCCGCGATCTCCGCGATGCTAAACCGCCGACCGCTGCGCTGTTTACAAAXCCGGATGCTGTTTTTGCATCGTT 600
 R T A I N A L A R N O B R G A I L D A D G T G D G C G C G L L D E P D R F P S I V
 GCGCAGGAGCGCTGGCTTTAGAAAATACGCTGTGCGGATGCTCTCTGAAATAGATCTGGAAGCTGCCCGCTGCCCGACCGCCATCTGTAAGAAGAACTGCAACGGAAACCC 720
 A Q E R L A P F R E N Y A V G R L P L E R A E T D P E L A A A A R R I V E R L O R N T
 TTTGCGATTGCGGCTGGCTGATGCTGCGCCACTAACGAAGCGCTGCTGGTGAATCGCCCTCTCTCTGCGCACCATGAGCAAAATTTTGTGAACCGCCCGCGCTGACCGTCCA 840
 L S I V G G R D V P T T E G V L G E I A L S L P R I E Q I F V N A P A G M R P
 CGCGATATGCAAGCGCTGTTTATCGCCCGCCCGCATGAAAGACCTCTGGAAGCGCAAAAGACTCTACGCTGTAGAGCTGCGAATCTGCTGCGAACAATCTATAAGCGCTGCTA 960
 R D N E R L P A R Y V I T R D E L I T C A S E V G I W D Y Q F D E V R K E R V G
 TCCCGACCGATCTGCCCGCTTTATCTGCATCTCCCGCCTGCTGCGAATGCGCAATTCGCTTCCACCAACCGCTTCCCATTAACCGGTACC CGCTGCCGCGCGCCGCAACCGT 1080
 C R R I C R V L S G S C C P A S C M A Z C L F P R P T N Y T P W R P W M A R
 TCCGCTATGCGCCCAATGCGAAATCAACACCACTACCGGTAAACCGCAATGGCGCTGCGCTACTATAAATTCAGACAGCGCTATCTCGTCAAGCGCCGCGCGCTGACCG 1200
 S A I W R I T V K S T P S P V T A N G R A R T Y K P O T P L I P D L M D A C H L D
 GTCACAAACCGCGCTGCTGCTCCAGTCTGATGATGATAACAGCTGCAAGCTCTGCTGCGACCGCGAATGCGATATCAACCGCCCAATGCGCTTATAGTACCAGCCCGCTGGCAGAAC 1320
 V H E T G E S D E S L M D H L L L L L G G M D I R A R N L V P P A W Q H
 CCGGATGTGACCGCGACTGCTGCTCTTCTTAACTCCATGATGAGCGCTGCGATGCG 1440
 P D M D P E L A R F P D F S N E M K P M D G P A C I V M S D G R P A C H L D
 CTAACCGCTGCGCGCGCGCT 1560
 R N O L R P A R Y V I T R D E L I T C A S E V G I W D Y Q F D E V R K E R V G
 CCAGCGCACTGATGCTTATGCGACCGCGCT 1680
 P G L E L N R A T D R S G R I L E S A R T D D D L E R S R F Y K E R S V R E K
 CTGTACCGCTTGAAGATCTGCGCAAGCTGCGGTAGCG 1800
 L V P F E D L P D E R L D D G S R E L D G D D L A S Y Q K Q F N Y S A R E L D S V I
 CCGCTACTGCGCGAAACCGTACGCAAGCGCT 1920
 R V L G E R G O G E A V G S R G D D D T P P A V L L S O P A I Y D V Y P Q O F A Q
 GTGACTAAGCGCGCAATGCG 2040
 V T H P I D P L R E A R M S L A T S I G R E N E V F C E T E K O A N B L S F
 AAATCCCGATCTGCT 2160
 K S P I L L Y L S D P F E G L P T N H E R E Y R A D T L D I T P D V Y T T L E A T
 GTCARAAGCTGTGCGCAAAACCGCAAAAGCT 2280
 V K E L C D R A A R K N V S G T V L L V L S D R N I A K R L D I P V P A P N A V G
 CGGATCCAGACCGCT 2400
 A I T O T R V L D G S L R C D A N I V E T A T A S A R D P H P F A V L L D G F G A T
 ATTATCTCACTGCT 2520
 I T Y P L Y L A Y E T L G R L V D T N A I A K D Y R T V M L N Y R N G I N K G L Y K
 ATCATCTCAAAATGCGCAATCCCACTGCT 2640
 I N S K M G I S T I A S Y R S E R L P A V G L N D D V G L C F G G A V S R I
 GGTGCAAGCTTTGAAGCTTCCAGCAGGATCTGCT 2760
 G A S P F Q D L L S K R A W L R X P I S Q G G L X R Y V H C G C E Y
 CAGCGCTACAGCG 2880
 M A N P D V D V T L Q O A V O S G E Y S D I O E Y A K L V N E R P A T T L R N
 CTGCTGCCAATACCG 3000
 L L A I T P G E N A V H I A D V E P A S E L L F R P R D T A A M S I G A L S F E A
 CAGGACCGCTGCG 3120
 M E A L A E A M S I C G H S N S G E G G E D P A R Y G G T N R V S R I K Q V A S
 GTCGCTTGGGCTACTCGCGCGTATCTGCTCAATGCCGCT 3240
 C R F G V T P A Y L V N A D V I O I R V A O G A X P G E G G L F G D R V T P Y
 ATCCGCAAGCT 3360
 I A K L R Y S V P G V T L Q O A V O S G E Y S D I O E Y A K L V N E R P A T T L R N
 AAAGCGATCTCCG 3480
 K A I S B V L K V S E P G V G T I A T G V A R A Y A D L I T I A G D D G C T G A
 AGTCGCTTTCATCGCTGAATACCG 3600
 S P A L S B V K Y A G G C P T W L G L V E T T O Q A L V A N G L R H X I A R L O V D G G C
 CTGAALCGCGCT 3720
 L R T G V D I I K A A I L G A E S F G F G T G P M V A L G C R Y L R I C H L M N
 TCCGCAACCGCTGACCAACTAGGATGCAAACTCGCTAAGAACCACTATCACCGCTGCCATCAAGCGTGACGAATFACTTGAATTAAGCGCGTGAACCGCGCGCGCGCGCG 3840
 C A T G V A T Q D X L R N N Y H T F L P P R V T N Y H F I E R I A R T E T R E L M A
 CAGCTTGGCGTAACAGCT 3960
 Q L C V T R L V D L I G R T D L L K E L D G F T A K R O O K L G L S R L L E T A E
 CCGCATCCAGCTAAGCGCTCTACTGCGCCGAAACAAACCGCGCTTTCATAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4080
 P H P G R A L Y T C T E M N P P F D N G L N L M A Q L L D Q A R P F V D E R G S T
 TTTGCTTGCATATGCGCAACCG 4200
 F M P D I R N T D R S V G S L S G Y I A O D A R R S G C L A A D P C T A A A G C T A C T T C A A C G G C
 ACCGACCG 4320
 T A G O S F G V M N A G G V E L Y L T G D A N D Y V G R C H A C G L I A I R P P

GTTGGTCCGCGCTCCGCAAGCCATGAAGCAACGATTCGTCGCTGTATGCGCCGACGGTGGTGGTGTATGCGCCAGCCGCGCGGCGTCAAGCTTCCGCGCTGCTAAC 4440
 VGEAPRREBEASIIIGWITCTLYGATGCGRLYAAARARERFQVNM
 TCCGGTGTATCACCCTGGTACAGGACATTCCGGACAAACCGTTGTAATATGACCGGCTGGTATCGCTGCAATTCGGGTAAACCGCGGCTTAAGTTCGGTGGGGCAGCAGCCGGT 4560
 SGAITVVEGIGDNGCEYMTGGIVCILGKRTGVNFGAGCMYGG
 TCCGCTACGTTCCGATGAAAGCGGATTTCCGCAACCGCTTAACCCGGAACTGGCCAGGCTTAAAGCGCTGACGCTCCGGATCATGAAGAGCATCGCTGGCTTTATACC 4680
 FATVYLDSESGDPRRERVNPELVLVLDALALAIHEENLRLGLIT
 GAGCATGTCAGCATACCGGCTCTCAGCCGCTGGAAGATTCCGGCAACTTCAACCTTCCGCCACTAAATTTCCGCTGGTAAACCTAAGTCCAGTGATGTAAGAACGCTCGGCT 4800
 ENVOETGSOEGEISLAAHWBTFATXFPALVXFPARLVDKALLG
 CACCGTAGTCCAGCGAGTTCGCGTCCAGCGCCAGTAAAGGGTAGCAACAATGAGTCAGAATGTTATCAATTTACGCTGCGCCGCTGATCCGCCAAAGAACCGCGT 4920
 NRSRBAAGAEELRVQAQEND [M] QN VY Q F I D L Q R V D P P K K P L
 AAGATCCGCAAAATTGAGTTTGTGAATTTACGAGCCGTTTCCGAAGCCAGGCGAAAGCCAGCGCTGACCGCTGCTGCTGGCGGAAACCCATACCTGCGAGTGGAAATGCCCGGTA 5040
 XIRKIEFVYIYEPFBEARFRRLTAACRAANPYCEMKCFV
 CACAACACTCCCGAAGCTGAGCTGCCAACGAGGGCGGTATTTGAAAGCGGCAACTGTCGCCACGACCAACACCCCTGCCGCGAAGTTCCGGACGAGCTGCCCGCAAGAC 5160
 HWYIFMMLKLLANEGRIPEAAEELSEHQNTLTPLEVCGRVCPOD
 CGTCTGCGAAGCTTCCGCTCAGCTGAGTGGTGGCGGCTGACCTCCGCAACATTAAGCGCTATATCAACGATAAAGCGCTCGAGATGGCGTCCGCGGATATCTCGT 5280
 RLCEGCTCTLMDSEDFGAVTIGHIERIYIRDKAFENMGWMPDHDHG
 GTGAACAGACCGGTAANAAGTGGGATATCCGCGCAGCCCGGAGGCTCCGCGATGCTCGCGGATGCTCGACCGCTAACCGGCTAAAAGCGGTTGCTTCCAGCGCTCATCAAGAAAT 5400
 VRTQKAGKAAVYAIIGAGFAGLACADVLTNRNGVXAVCFDRHPEI
 GCGCGCTGCTGACTCCGTTATCCGCGCTTCAAGCTGGAAGAGAGGTAATGAGCGCTCCGCTGAAATCTTCCAGCCGATCGGTTGCAATCAAACCTCAATACCGAAGTGGCGCG 5520
 GGLLTFGIFAFKLEEVNTRRREIFTGNGIEFKLNTYEVGR
 GAGTCACGCTGGACATCTGCTGATGATTACGATCCGCTTCTTCCGCTCCGCACTATCATCTCAATCCGCGCGCGCGGCAAAACGAAGACCGCGTACCGCAGCCGCTG 5640
 DVQLDDLLLSDYDAVFLGVGTYSMRGCGLENEDEDGCTORC
 COTTCCTATCCGCAACGCAACGACTTAAAGCGCTTGGTGAARCCCGCAGCAACCGTTTCGTCAGCATGGAAGCCAAAGCCGCTGGCTCCCTTCCGCGCTGGCAGACTCGATGGAC 5760
 RSESTPFPHSLRTRDRDFVSEHEZKRVVVLGGCGDTAHD
 TCCGCTGCTACCTCCGCGCAGGCGGCAAGCAGCTTACCTGCTGCTATCGTGTGATGAAGAAATGCGCGGCTCCCGCGCGGAAATGAAAACCGCCGGAAAGCGGCTAGAC 5880
 CVRTSTVROQANDVTCAYRDRDENRPNPGBRERVDKXARAEZGVE
 TTCAAATCAACGCTCCAGCGCTGATCAAGGTCAAGCGTAAAGTCAGCGGCTGAAAATGGTCCGATCCGAAATGCGCGCAAGCGCTGCGCCGCGGAGAT 6000
 KFNVOQPLGIEVNGHGGKLVGKVMVRETEENGEFPAKABRFGD
 CGTTCAGGTTCCGAACATATCGAAGCCGAGATCGCGTGCATGCGGCTTGGTTTCGCCACACAACATGGAATGCGTCCGCAAAACACAGCGCTCGAGCTGGATTACAAGCCCGCATC 6120
 RCRPRTYRTGRGCDHGVMWFRPHNMELAKHEVELEDSQGR I
 ATCGCCGGGAAGCCAGCAGCGCTTCCAGACAGCAGCCGAAATCTTTCGCTGGATATCGTCCGCTGGTTCGATGGTGGTGAACCGCTATTCGCGAAGCGCTGAAGCGC 6240
 IAPRGS DNA FDT SMPRI F AGGDIVRGSDLVVTAIAZGRKA
 GCAGCGGTATGAACTGGCTGGAAGTTAAGCGAGGTAACAATGAATC
 A D G I M N W L E V E N D

Fig. 2. Nucleotide sequence of a 6.3-kb *HpaI-EcoRI* *E. coli* K-12 DNA fragment cloned in pRSP20. The sequence comprises the coding regions for the large and small GOGAT subunits and their 5'- and 3'-flanking regions (see Fig. 1). The DNA region analyzed also contains the identified transcriptional start point (wavy arrow, nt position 178) and its corresponding promoter with the -10 and -35 regions overlined. A palindromic sequence centered 30 bp after the transcriptional start point is indicated by facing arrows. Partially overlapping the palindromic sequence, there is a sequence similar to the *boxA* antiterminator consensus sequence (see Fig. 6). Hypothetic RBS are overscored and indicated by S.D. The large subunit gene has two possible translation initiation methionine residues (boxed, nt positions 301 and 394) with corresponding RBS, separated by 30 aa residues. The mature N termini (nt positions 427 and 4861) residues for the two subunits are circled. Stop codons for both subunits are indicated by 'END'.

after posttranslational processing. In this enzyme, an 11-aa peptide leader is removed and a cysteine residue is exposed as the N terminus. Furthermore, this residue has been involved in this enzyme, as part of the glutamine binding site (Makaroff et al., 1983; Vollmer et al., 1983).

Based on the same protein sequence results, the mature small GOGAT subunit was shown to start with a serine residue (Fig. 2). This result suggests that the preceding methionine residue (nt position 4846, Fig. 2) is the translational start site for this subunit and that this residue is posttranslationally removed.

The reported estimated M_r s for the *E. coli* W GOGAT subunits are 135000 and 53000 (Miller

and Stadman, 1972). The M_r for the *E. coli* K-12 small subunit deduced from the nucleotide sequence is in good agreement with the reported one. However, in *E. coli* K-12 the first ORF codes for a protein of larger M_r than the one reported for the *E. coli* W large subunit. We confirmed the M_r obtained from the nucleotide sequence by comparing the mobility of the large GOGAT subunit with other proteins of known M_r in an SDS-PA gel. Therefore, we believe that the M_r for the *E. coli* K-12 large GOGAT subunit, as predicted by the nucleotide sequence, is 166208, similar to that for the large GOGAT subunit of *Klebsiella aerogenes* (Trotta et al., 1974).

The analysis of the nucleotide sequence located at the 5' position of the gene that codes for the large

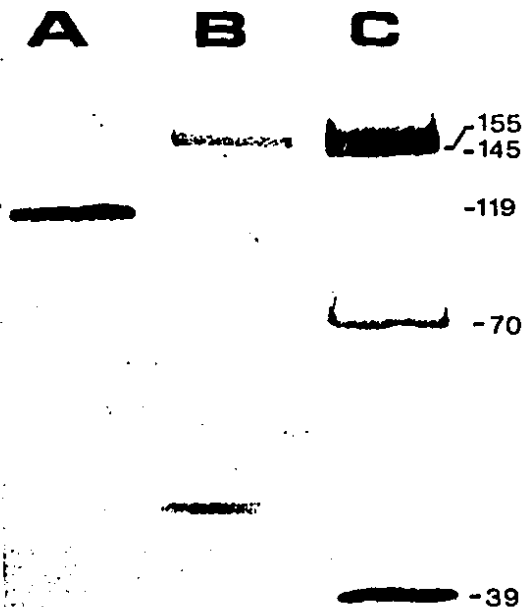


Fig. 3. M_r -value determination of the GOGAT subunits by SDS-PA gel electrophoresis. Two or three μ g of each sample were loaded into a 0.1% SDS-11% polyacrylamide gel. Lanes: A, *E. coli* β -galactosidase; B, purified GOGAT enzyme; C, *E. coli* RNA polymerase. Sizes are indicated on the right margin in kDa.

subunit revealed two putative promoters based on consensus parameters (Hawley and McClure, 1983). The extension experiments localized the transcriptional start point at nt position 178. Inspection of the nucleotide sequence upstream from this start point revealed that the functional promoter, for the conditions used, is one of the two putative promoters identified by consensus comparison. This promoter has a -35 sequence which is identical in five out of six positions to the consensus sequence. The -10 sequence has three identities to the consensus sequence. The distance between the proposed -35 and -10 regions is 17 bp which is the optimum for promoter activity (Hawley and McClure, 1983). A DNA region with a high A + T content (65%) is also present preceding this promoter. A + T-rich regions have been observed preceding various prokaryotic

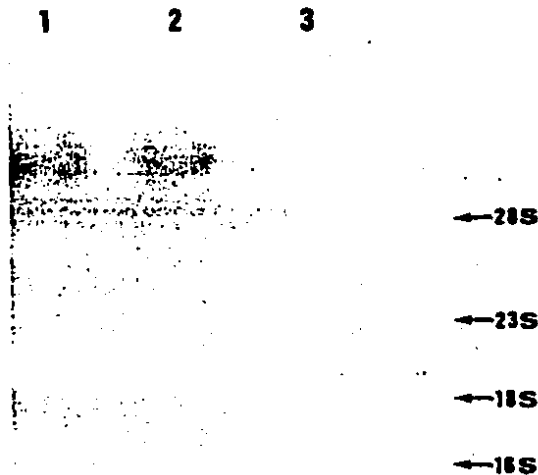


Fig. 4. Northern blot analysis of GOGAT transcripts. RNA was isolated as described in MATERIALS AND METHODS, section f. Samples of total RNA (10 μ g) from strains CS520, PA340 and PA340[pRSP20] were subjected to electrophoresis in a 1.5% agarose 2.2 M formaldehyde gel, RNA was transferred to nitrocellulose and hybridized with an *EcoRI-EcoRI* probe corresponding to the small subunit gene (nt positions 5495-6289, Fig. 1). Hybridization using a *BglII-PvuII* probe located in the large subunit gene (nt positions 1697-4823, Fig. 1) gave the same results (not shown). The probes were labeled with [α - 32 P]dCTP by nick translation. Lanes: 1, RNA from strain CS520 grown in NN medium with 15 mM NH_4Cl ; 2, RNA from strain CS520 grown in NN medium with 0.5 mM NH_4Cl ; 3, RNA from strain PA340 grown in NN medium with 0.5 mM NH_4Cl . The same results as in CS520 were obtained using RNA from strain PA340[pRSP20] (not shown). RNA site markers used were rat 28S (5.5 kb) and 18S (2.1 kb) rRNA and *E. coli* 23S (3.1 kb) and 16S (1.5 kb) rRNA. The bands above 28S rRNA in lanes 1 and 2, are specific GOGAT transcripts of approx. 7.3 kb.

promoters (Vollenweider et al., 1979). We also identified a putative CRP-cAMP binding site overlapping the -35 region of this functional promoter (deCrombrughe et al., 1984). This finding agrees well with earlier reports which showed that the production of GOGAT is negatively controlled by CRP-cAMP (Prusiner et al., 1972). It is important to notice that in some genes, negatively controlled by CRP-cAMP, this binding domain similarly overlaps the -35 promoter region (deCrombrughe et al.,

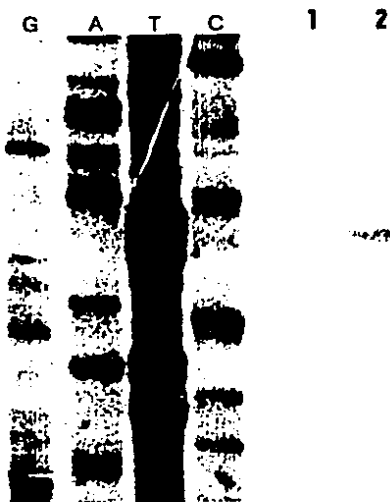


Fig. 5. Transcriptional start point mapping. The transcriptional start point was mapped using a modification of the protocol for primer extension reported by León et al. (1985). 50 μ g of total RNA from strain PA340 or PA340[pRSP20] was mixed with 1 pmol of [5'- 32 P]oligodeoxynucleotide primer (30-mer) which is complementary to the region at nt positions 473-502 (Fig. 2). The mixture was denatured at 100°C for 5 min in 8.7 mM Tris · HCl (pH 8.3)-0.35 mM EDTA and chilled on dry ice. Primer-RNA hybridizations were incubated at 43°C for 3 h. Synthesis of cDNA was carried out in 52 mM Tris · HCl (pH 8.3), 10.4 mM MgCl₂, 4.2 mM dithiothreitol, 1 mM each dATP, dCTP, dGTP, and dTTP, and 30 units of reverse transcriptase in a final volume of 30 μ l and incubated for 1 h at 43°C. RNA was degraded with 1 μ l of an RNase A solution (1 mg/ml), for 1 h at 37°C. After phenol extraction and ethanol precipitation, the cDNA was resuspended in 3 μ l of water and 6 μ l of stop mix (95% formamide, 0.02% xylene cyanol and 0.02% bromophenol blue). The reverse transcribed material was electrophoresed through a 5% PA-8 M urea gel. The extended product is indicated by an arrow. The extension experiments were performed using RNA from the *E. coli* K-12 strains PA340 (lane 1), and PA340[pRSP20] (lane 2). The nucleotide sequence products of M13mp18 phage DNA (G, A, T, C), were used as *M_r* markers.

1984). In addition, we also identified a region of dyad symmetry, suggesting an 'operator-like' structure (nt positions 197-215, Fig. 2). We would like to speculate that this structure could influence the transcription of the GOGAT genes under certain metabolic conditions.

An additional feature found by nucleotide sequence analysis was the presence of a putative anti-terminator *boxA*-like sequence. This sequence is part of the GOGAT leader mRNA transcript. In *E. coli* this sequence has been found at the same position in several polycistronic transcripts, e.g. the *rrnB* operon, and it has been conclusively demonstrated, for phage *N-nut* system, that it is part of an anti-termination mechanism (Olson et al., 1984; Hasan and Szybalski, 1986a,b). As shown in Fig. 6, the similarity between the *rrnB* and GOGAT mRNA leaders around the *boxA* flanking regions, is high. However, at the present time, we do not have any physiological support for an antitermination event in GOGAT. Nevertheless, such a mechanism could be involved as part of a cellular strategy to avoid fortuitous and premature transcription termination events (Olson et al., 1984; Szybalski et al., 1987).

The analysis of the intercistronic DNA region, and also of the end of the large GOGAT gene, revealed other interesting features. The observation that the TAA termination codon for the large subunit overlaps the potential RBS of the small GOGAT subunit, and that there is a close proximity between the translational termination codon of the large subunit and the initiation codon of the small subunit, suggests that both genes could be translationally coupled. Translational coupling in *E. coli* has been previously reported for various operons such as the tryptophan (Oppenheim and Yanofsky, 1980), galactose (Schumperli et al., 1982), and ribosomal protein operons (Yates and Nomura, 1981).

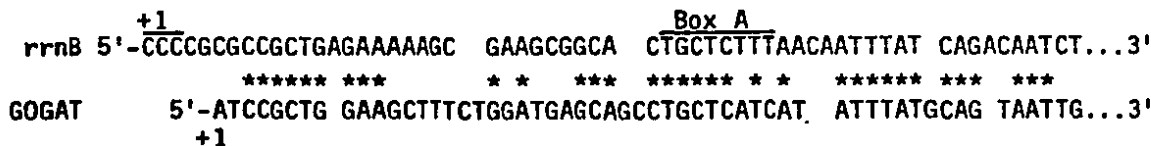


Fig. 6. Comparison between the transcript leader region of the *E. coli rrnB* operon (upper sequence) and the untranslated leader region of the *E. coli* GOGAT mRNA (lower sequence). Identities are indicated by asterisks. To maximize alignment, gaps were introduced. The transcriptional start points are indicated by +1; the *boxA* antiterminator-related sequence is indicated above the *rrnB* leader. The *boxA* for phage λ is CGCTCTTA (Olson et al., 1984).

Although the translational coupling mechanism of two adjacent genes is poorly understood at present, it is thought that the close proximity of overlapping translational termination and initiation codons ensures the coordinated expression of genes with related functions (Oppenheim and Yanofsky, 1980).

It is also important to notice that the distance between the termination codon TAA for the large subunit and the translational start point of the small subunit is 12 nt. One could speculate that if a point mutation abolished this stop codon, it might be possible to obtain a fusion protein of approx. 218 kDa. This observation is interesting from an evolutionary point of view, since in *N. crassa* and in some plants it has been shown that the GOGAT enzyme consists of a single polypeptide of approx. 220 kDa (Hummelt and Mora, 1980).

ACKNOWLEDGEMENTS

We wish to thank Gilberto Mosqueda for his help in the purification of the GOGAT enzyme. The protein sequence determination was performed by Guillermo Ramírez in the Protein/DNA Sequence Synthesis Facility of the University of Wisconsin Biotechnology Center, Madison, WI (U.S.A.), supported by funds from the Public Health Service, National Institutes of Health (Shared Equipment grant S10-RR1684, National Cancer Institute continued support grant CA-07175 and the General Research Support grant to the University of Wisconsin Medical School) and from the University of Wisconsin Graduate School. We wish to thank Dr. Edmundo Calva who first developed some of the purification manipulations of the GOGAT enzyme at the University of Wisconsin Biotechnology Center. G. Gosset is a recipient of a fellowship from Consejo Nacional de Ciencia y Tecnología, México. This work was supported in part by Consejo Nacional de Ciencia y Tecnología, México, Donativo PCCBNAL-022584.

REFERENCES

Berberich, M.A.: A glutamate-dependent phenotype in *E. coli* K-12; the result of two mutations. *Biochem. Biophys. Res. Commun.* 47 (1972) 1498-1503.

- Clarke, L. and Carbon, J.: Biochemical construction and selection of hybrid plasmids containing specific segments of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* 72 (1975) 4361-4365.
- Covarrubias, A.A., Sanchez-Pescador, R., Osorio, A., Bolivar, F. and Bastarrachea, F.: ColE1 plasmids containing *Escherichia coli* genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* 3 (1980) 150-164.
- De Banzie, J.S., Steeg, E.W. and Lis, J.T.: Update for users of the Cornell sequence analysis package. *Nucl. Acids Res.* 12 (1984) 619-625.
- De Boer, H.A., Comstock, L.J., Hui, A., Wong, E. and Vasser, M.: Portable Shine-Dalgarno regions: nucleotides between the Shine-Dalgarno sequence and the start codon affect the translation efficiency. In Papas, T.S., Rosenberg, M. and Chirikjian, J.G. (Eds.), *Gene Amplification and Analysis*, Vol. 3. Elsevier, Amsterdam, 1983, pp. 103-116.
- deCrombrugge, B., Busby, S. and Buc, H.: Cyclic AMP receptor protein: role in transcription activation. *Science* 224 (1984) 831-838.
- Fristensky, B., Lis, J. and Wu, R.: Portable microcomputer software for nucleotide sequence analysis. *Nucl. Acids Res.* 10 (1982) 6451-6463.
- Garcarrubio, A., Lozoya, E., Covarrubias, A. and Bolivar, F.: Structural organization of the genes that encode two glutamate synthase subunits of *Escherichia coli*. *Gene* 26 (1983) 165-170.
- Gourse, R.L., De Boer, H.A. and Nomura, M.: DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, anti-termination. *Cell* 44 (1986) 197-205.
- 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.
- Hasan, N. and Szybalski, W.: Boundaries of the *nutL* anti-terminator of coliphage lambda and effects of mutations in the spacer region between *boxA* and *boxB*. *Gene* 50 (1986a) 87-96.
- Hasan, N. and Szybalski, W.: Effect of the promoter structure on the transcription antitermination function. *Gene* 50 (1986b) 97-100.
- Hawley, D.K. and McClure, W.C.: Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids Res.* 11 (1983) 2237-2255.
- Hui, A., Hayflick, J., Dinkespiel, K. and De Boer, H.A.: Mutagenesis of the three bases preceding the start codon of the β -galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO J.* 3 (1984) 623-629.
- Hummelt, G. and Mora, J.: NADH-dependent glutamate synthase and nitrogen metabolism in *Neurospora crassa*. *Biochem. Biophys. Res. Commun.* 92 (1980) 127-133.
- Leon, P., Romero, D., Garcarrubio, A., Bastarrachea, F. and Covarrubias, A.A.: Glutamine synthetase-constitutive mutation affecting the *glnALG* upstream promoter of *Escherichia coli*. *J. Bacteriol.* 164 (1985) 1032-1038.
- Lozoya, E., Sanchez-Pescador, R., Covarrubias, A.A., Vichido, I. and Bolivar, F.: Tight linkage of genes that encode the two

- glutamate synthase subunits of *Escherichia coli* K-12. *J. Bacteriol.* 144 (1980) 616-621.
- Makaroff, C.A., Zalkin, H., Switzer, R.L. and Vollmer, S.J.: Cloning of the *Bacillus subtilis* glutamine phosphoribosylpyrophosphate amidotransferase gene in *Escherichia coli*. *J. Biol. Chem.* 258 (1983) 10586-10593.
- Mantsala, P. and Zalkin, H.: Glutamate synthase properties of the glutamine-dependent activity. *J. Biol. Chem.* 251 (1976a) 3294-3299.
- Mantsala, P. and Zalkin, H.: Properties of apoglutamate synthase and comparison with glutamate dehydrogenase. *J. Biol. Chem.* 251 (1976b) 3300-3305.
- Mantsala, P. and Zalkin, H.: Active subunits of *Escherichia coli* glutamate synthase. *J. Bacteriol.* 126 (1976c) 539-541.
- Messing, J., Crea, R. and Seeburg, P.H.: A system for shotgun DNA sequencing. *Nucl. Acids Res.* 9 (1981) 309-321.
- Miller, E.R. and Stadman, E.R.: Glutamate synthase from *Escherichia coli*: an iron sulfide flavoprotein. *J. Biol. Chem.* 247 (1972) 7109-7419.
- Normark, S., Bergstrom, S., Edlund, T., Grundstrom, T., Jaurin, B., Lindberg, F.P. and Olsson, O.: Overlapping genes. *Annu. Rev. Genet.* 17 (1983) 499-525.
- Oliver, G., Valle, F., Rosetti, F., Gomez-Pedrozo, M., Santamaria, P., Gosset, G. and Bolivar, F.: A common precursor for the two subunits of the penicillin acylase from *Escherichia coli* ATCC11105. *Gene* 40 (1985) 9-14.
- Olson, E.R., Tomich, C.C. and Friedman, D.L.: The *nusA* recognition site. *J. Mol. Biol.* 180 (1984) 1053-1063.
- Oppenheim, D.F. and Yanofsky, C.: Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* 95 (1980) 785-795.
- Prusiner, 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.
- Rendina, A.R. and Orme-Johnson, W.H.: Glutamate synthase: on the kinetic mechanism of the enzyme from *Escherichia coli*. *W Biochemistry* 17 (1978) 5388-5393.
- Rosenfeld, S.A. and Brenchley, J.E.: Regulation of glutamate and glutamine biosynthesis. In Herrmann, K.M. and Somerville R.L. (Eds.), *Amino Acids: Biosynthesis and Genetic Regulation*. Addison-Wesley, Cambridge, MA, 1983, pp. 1-17.
- Sanchez-Pescador, R. and Urdea, M.S.: Use of unpurified synthetic deoxynucleotide primers for rapid dideoxynucleotide chain termination sequencing. *DNA* 3 (1984) 339-343.
- Sakamoto, N.A., Kotre, A.M. and Savageau, M.A.: Glutamate dehydrogenase from *Escherichia coli*: purification and properties. *J. Bacteriol.* 124 (1975) 775-783.
- Schumperli, D., McKenny, K., Sobieski, D.A. and Rosenberg, M.: Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* 30 (1982) 865-871.
- Stormo, G.D., Schneider, T.D. and Gold, L.M.: Characterization of translational initiation sites in *E. coli*. *Nucl. Acids Res.* 10 (1982) 2971-2995.
- Szybalski, W., Brown, A.L., Hasan, N., Podhajnska, A.J. and Somasekhar, G.: Modular structure of the *nut* and *gut* anti-terminators of transcription. Interactions between control elements of phage lambda and construction of novel regulatory circuits. In Reznikoff, W.S., Burgess, R.R., Dahlberg, J.E., Gross, C.A., Record, M.T. and Wickens, M.P. (Eds.), *RNA Polymerase and the Regulation of Transcription*. Elsevier, New York, 1987, pp. 381-390.
- Trotta, P.P., Platzer, K.E.B., Haschemeyer, R.H. and Meister, A.: Glutamine-binding subunit of glutamate synthase and partial reactions catalyzed by this glutamine amidotransferase. *Proc. Natl. Acad. Sci. USA* 71 (1974) 4607-4611.
- Tyler, B.: Regulation of the assimilation of nitrogen compounds. *Annu. Rev. Biochem.* 47 (1978) 1127-1162.
- Vollenweider, H.J., Fiandt, M. and Szybalski, W.: A relationship between DNA helix stability and recognition sites for RNA polymerase. *Science* 205 (1979) 508-511.
- Vollmer, S.J., Switzer, R.L., Hermodson, M.A., Bower, S.G. and Zalkin, H.: The glutamine-utilizing site of *Bacillus subtilis* glutamine phosphoribosylpyrophosphate amidotransferase. *J. Biol. Chem.* 258 (1983) 10582-10585.
- von Heijne, G.: A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* 14 (1986) 4683-4690.
- Yates, J.L. and Nomura, M.: Feedback regulation of ribosomal protein synthesis in *Escherichia coli*: localization of the mRNA target site for repressor action of ribosomal protein L1. *Cell* 24 (1981) 243-249.
- Young, F.S. and Furano, A.V.: Regulation of the synthesis of *E. coli* elongation factor Tu. *Cell* 24 (1981) 695-706.

Communicated by Z. Hradečná.



PROTEIN IDENTIFICATION RESOURCE

NATIONAL BIOMEDICAL RESEARCH FOUNDATION
Georgetown University Medical Center
3900 Reservoir Road, N.W.
Washington, D.C. 20007

Tel (202) 687-2121

May 4, 1988

Dr. Francisco Bolivar
Departamento de Biología Molecular
Centro de Investigación sobre
Ingeniería Genética y Biotecnología
Universidad Nacional Autónoma de México
Apartado Postal 510-3, Cuernavaca, Morelos
C.P. 62270 México

Dear Dr. Bolivar,

Enclosed is your manuscript with the comments of the reviewers. The reviewers consider this paper suitable for publication in Protein Sequences & Data Analysis after revision. I hope you find their suggestions helpful and look forward to receiving the revised manuscript.

Please note the format of references as shown in the enclosed "Instructions to Authors".

Sincerely yours,

Winona C. Barker

**Amino acid sequence analysis of the glutamate synthase enzyme
from Escherichia coli K-12**

**Guillermo Gosset, Enrique Merino, Felix Recillas, Guillermo
Oliver+, Baltazar Becerril and Francisco Bolivar**

**Departamento de Biología Molecular, Centro de Investigación sobre
Ingeniería Genética y Biotecnología, Universidad Nacional
Autónoma de México, Apartado Postal 510-3, Cuernavaca, Morelos
C.P. 62270 (México)**

Offprint requests to: Francisco Bolivar.

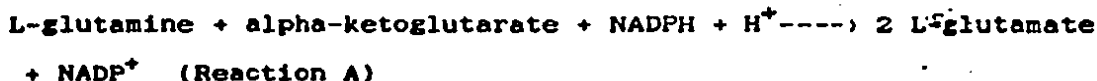
**+ Present address: Department of Biological Chemistry, UCLA
School of Medicine, Los Angeles, CA 90024 (USA)**

**Abbreviations: GOGAT, glutamate synthase; GDH, glutamate
dehydrogenase; SD, standard deviation units.**

Abstract. The amino acid sequence for the two subunits of the glutamate synthase of Escherichia coli K-12 was compared to the protein sequences compiled in the National Biomedical Research Foundation databank. Similarities were detected between the small glutamate synthase subunit and three members of the flavin-containing pyridine nucleotide-disulphide oxidoreductase superfamily, and also with three members of a lactate dehydrogenase family. Two segments in this glutamate synthase subunit showed similarity to regions previously proposed as part of dinucleotide binding sites in some members of these two families. Similarity can be extended if the predicted secondary structure is considered. Based on these data, residues 148-260 and 289-409 in the small GOGAT subunit are proposed as dinucleotide binding regions.

Comparison of the amino acid sequence of the large glutamate synthase subunit with the glutamine phosphoribosylamine:pyrophosphate phosphoribosyltransferases of B. subtilis and E. coli revealed a significant similarity between the amino termini of these three enzymes. In these last two amidotransferases, the glutamine-binding site has been located in their amino-terminal region. The comparison with a second group of glutamine amidotransferases did not show any significant global similarity with the large glutamate synthase subunit. However, this polypeptide contains a small segment that shares similarity with a 13 amino acid segment proposed as part of the glutamine-binding site in this second group of amidotransferases.

Glutamate synthase (GOGAT) (L-glutamine: alpha-ketoglutarate aminotransferase (reduced NADPH⁺ oxidizing), EC 2.6.1.53) catalyzes the reaction:



It has been demonstrated that GOGAT also catalyzes an ammonia-dependent synthesis of glutamate:



However, this activity represents only 5-7% of the glutamine-dependent activity (Reaction A). Apparently, this reaction is very similar to the one catalyzed by the NADPH-dependent glutamate dehydrogenase (GDH) [1].

GOGAT is present in several biological taxa, including bacteria [2], algae [3], fungi [4], and higher plants [5]. In enterobacteria, growth under conditions of nitrogen limitation is accomplished by the coupling of Reaction A to the ATP-dependent synthesis of glutamine from ammonia and glutamate catalyzed by glutamine synthetase [2].

The GOGAT enzyme is a complex system. In *E. coli* K-12 this

enzyme is composed of two subunits with Mr of approximately 166 kDa (1514 aa) and 52 kDa (417 aa) for the large and small subunits respectively [6]. These subunits are isolated in vivo as an $\alpha_4\beta_4$ subunit structure of approximately 800 kDa [7].

GOGAT is an iron-sulfur flavoprotein [8]. Sedimentation velocity studies suggested that the iron-sulfide and flavin prosthetic groups are located on the large subunit [7]. These groups are required for the glutamine-dependent glutamate synthesis [9].

Biochemical studies using a chloro-ketone analog of glutamine, have demonstrated that this reagent binds only to a cysteine residue in the large subunit, causing the inhibition of the glutamine-dependent activity of the enzyme (Reaction A), while its ammonia-dependent activity (Reaction B), remains unaltered [7, 9].

Previous reports have established similarities and differences between GOGAT and other glutamine amidotransferases [1, 7, 9, 10]. Like these enzymes, GOGAT has a glutaminase activity, and as mentioned above, it can utilize ammonia instead of glutamine for the synthesis of glutamic acid [9]. Main differences are the presence of iron-sulfur and flavin groups and the binding of glutamine to the large subunit in GOGAT.

We have recently determined the complete nucleotide sequence of the GOGAT structural genes from Escherichia coli K-12 [6]. The deduced amino acid sequences for both subunits allowed a systematic comparison study with protein sequences compiled in the National Biomedical Research Foundation databank. This is, as

far as we know, the first reported comparison of the amino acid sequence of a glutamate synthase enzyme with a protein sequence databank. This study has allowed us to assign putative functions to certain specific segments in both GOGAT subunits, and also to identify possible relationships with some proteins whose sequences are compiled in the databank.

Materials and methods

Protein sequence comparisons were carried out using a VAX 11/780 computer, with the VAX/VMS system. The amino acid sequence of each GOGAT subunit was compared to the National Biomedical Research Foundation databank release 10.0, initially using the FASTP program [11]. In order to find further relationships, the sequences with the highest scores were also analyzed with the RELATE or ALIGN programs [12] using the mutational data matrix [13]. To determine the statistical significance of these studies, 50 random sequences, with the same amino acid composition as one of the sequences studied, were generated and compared. Secondary structure analysis was performed using the method proposed by Chou and Fasman [14, 15].

Results and discussion

Databank comparisons

The amino acid sequence of each GOGAT subunit was compared with the National Biomedical Research Foundation databank, using the FASTP program [11]. We further analyzed the proteins exhibiting the best 30 scores with respect to each of the two subunits. Some of the protein segments chosen by the FASTP program were clustered at specific GOGAT regions, suggesting the existence of specific segments that may have a common function among these proteins.

Specific protein domains from several dinucleotide-binding enzymes showed similarity to two specific regions located in the small GOGAT subunit. Fig. 1 shows the location of these regions, designated as C and E (Fig. 1).

General analysis of similarities

Further analysis was carried out with the RELATE program since the FASTP program finds only the region with the highest similarity between two protein sequences, and the statistical significance of such similarities is not obtained. We also included in this analysis the sequences of several proteins that share some functions with GOGAT, although they were not selected by the previous FASTP search.

Analysis of similarities between the small GOGAT subunit and proteins obtained from the FASTP analysis

A sequence comparison was performed using the ALIGN and RELATE programs. The ALIGN analysis did not reveal any global statistically significant similarity between the small GOGAT subunit and the various enzymes obtained with the FASTP analysis.

However, from the RELATE analysis, three members of the flavin-containing pyridine nucleotide-disulphide oxidoreductases superfamily [16] and three proteins from a lactate dehydrogenase family [17, 18, 19], showed statistically significant scores (score > 3 SD) when compared with the small GOGAT subunit (see Table I). *E. coli* GDH, which was not found with the FASTP program, was also analyzed since it catalyzes a reaction similar to the one identified for the small GOGAT subunit (Reaction B) [1]. This last comparison produced a score of 2.23 SD. The parameters used with the RELATE program allowed the comparison of specific segments from these proteins. Thus, the scores obtained indicate local not global similarities.

This RELATE analysis also revealed that several segments in the enzymes that obtained scores higher than 3 SD, were similar to two separated regions in the small GOGAT subunit. One of these regions, position 148-178 (segment "c"), is located in cluster C and was originally detected with the FASTP program. The other region, not previously detected, was localized at position 289-319 (segment "d"), cluster D, Fig. 1. From this analysis we also found that the two proteins represented in cluster E have other segments that share similarity to regions present in clusters C and D, as indicated in Fig. 1.

Table II shows an alignment of segments (identified as part of the dinucleotide-binding domains) of the enzymes from clusters C

and D , with segments "c" and "d" from the small GOGAT subunit.

The primary sequence similarity led us to search for secondary structure similarities between these regions, since several dinucleotide-binding domains possess common structural features. [20, 21]. We used the secondary structure prediction method proposed by Chou and Fasman [14, 15] to generate the prediction profile shown on Fig. 2. Both GOGAT segments "c" and "d" display a potential beta-alpha-beta structure, which is present as part of the dinucleotide binding domain in several proteins. Following this region, the prediction profile shows an alternation of alpha and beta segments and no regions of random coil. These predicted structural features are similar to those described by Rossmann et al. [20] and Wootton [21], as part of a dinucleotide binding fold. These data indicated that in the small GOGAT subunit, segments "c" and "d" share similarity between them. This is the reason why all protein segments present in cluster C are also members of cluster D.

Furthermore, certain features in the beta-alpha-beta structural motif of some dinucleotide binding enzymes [22], are present in these two GOGAT segments (see Table II):

1.-The conserved sequence Gly-X-Gly-X-X-Gly is found in segment "c", position 153-158 and partially conserved in segment "d", position 294-299.

2.-In segment "c", four of the six hydrophobic residues expected are present in the positions equivalent to the hydrophobic core of the beta-alpha-beta unit. In segment "d", four hydrophobic residues are found in the equivalent positions. The expected

positions for these hydrophobic residues are indicated in Table II below the aligned sequences with a letter "H".

3.-At the C-terminus of the second beta sheet region, a negatively charged residue is expected in the case of the NAD or FAD-binding beta-alpha-beta units. An aspartic acid residue is found in segment "c" (position 176).

4.-An hydrophilic residue, at the beginning of the predicted first beta sheet region, is present in segment "c" (lysine, position 148) and in "d" (arginine, position 289).

In the case of several dehydrogenases, an arginine residue located immediately after an acidic one at the carboxyl terminus of the second beta sheet, has been implicated in the interaction with the phosphate group of the NADP cofactor, [16, 23]. An arginine residue located immediately after an aspartic acid residue (position 178) is present in the second beta sheet in segment "c" in the small GOGAT subunit (Table II).

These considerations lead us to propose that segments "c" and "d" have similar sequence and structural features to the beta-alpha-beta unit that has been defined by three-dimensional structure analysis as a component of the dinucleotide-binding domain [20, 22]. Segment "c" shows a higher degree of conservation for these features so it is the most probable dinucleotide-binding region in the small GOGAT subunit.

A common origin has been proposed for the mononucleotide binding fold present in various proteins. A duplication of this structural domain could have generated the dinucleotide binding motif [24]. It has also been proposed that certain proteins have been assembled by the fusion of structural modules with specific

functions [25]. The dinucleotide binding regions in the glutathione reductase enzyme, a member of the flavin-containing pyridine nucleotide-disulphide oxidoreductase family, were apparently originated by gene duplication. Divergence between these two sites could allow the recognition of FAD by one site and of NADP by the other [26]. Although the small GOGAT subunit did not yield a statistically significant score when it was compared to the glutathione reductase (2.42 SD), it did with three other members of the flavin-containing pyridine nucleotide-disulphide oxidoreductase superfamily (Table I). Two of them, lipoamide dehydrogenase and mercuric reductase, contain two dinucleotide binding folds. The present data could suggest gene duplication or fusion as the cause for the two dinucleotide binding regions in the small GOGAT subunit. Another possible origin could be divergence from a protein that already possesses two dinucleotide binding sites, perhaps an ancestral protein related to the enzymes presented in Table I. This last hypothesis is supported by the fact that all those enzymes, including the small GOGAT subunit, have a Mr around 51 kDa. Also, the sequence similarity between segments "c" and "d", is not significantly higher than the similarity between either of these two regions and any of the beta-alpha-beta folds presented in Table II (data not shown).

The existence of two possible dinucleotide-binding sites has also been proposed for the *N. crassa* and bovine GDHs [21]. Both domains of *N. crassa* GDH seem to be important for dinucleotide binding, although only one exhibits the general characteristics

of such a site [22]. We have previously [27] reported similarity between GDH amino acid segments from different sources including the two proposed nucleotide binding domains for N. crassa and bovine GDHs. If the small GOGAT subunit employs a similar mechanism for dinucleotide binding, then both regions could be required, even though the second domain (segment "d") is less conserved than the first one (segment "c").

Furthermore, Rendina and Orme-Johnson [28] have reported an in vitro activation of GOGAT by NADP^+ at low levels of alpha-ketoglutarate. This activation results from lowering the K_m for alpha-ketoglutarate and, as proposed by the authors, might involve the binding of NADP^+ to a different site from the one for NADPH. Therefore, the less conserved dinucleotide binding site in the small GOGAT subunit could play a role in the modulation of the enzyme activity.

Based on similarities in catalytic properties and Mr between GDH and the small GOGAT subunit from E. coli, there has been speculation about a relationship between them. However, Mäntsälä and Zalkin [1], using microcomplement fixation techniques and double antibody precipitation, were not able to detect relatedness between these enzymes. A comparison of the primary sequence of these proteins with the RELATE program, using the mutational data matrix, produced a score (2.23 SD) which does not clearly support a possible relationship. However, it is important to remark, that the NADPH-binding interaction in the E. coli GDH probably involves two regions (unpublished observations), and the small GOGAT subunit has two possible dinucleotide-binding regions. This similarity might be the reason why the RELATE score

is higher than most of the scores obtained from a random comparison. Nevertheless, the similarity is apparently not sufficient to be detected by immunological methods.

Analysis of similarities between the large GOGAT subunit and the glutamine binding subunit of several glutamine amidotransferases

A comparison of the protein sequences obtained from the FASTP search with the large GOGAT subunit was made using the RELATE and ALIGN programs. This analysis did not yield any global statistically significant score for any case (data not shown).

Nevertheless, the large GOGAT subunit and the glutamine binding subunit of several glutamine amidotransferases (that did not appear in the FASTP analysis) have certain common characteristics: the existence of a glutaminase activity and the presence of a cysteine residue reactive to glutamine analogs [29]. In order to find out if these common features could be reflected by the presence of similar amino acid sequences, a comparison among them was performed using the RELATE program. The resulting scores (data not shown) indicate that the glutamine amidotransferases studied can be divided in two groups, according to sequence similarity, as it has been previously reported [29]. In this work we use the terms Group "I" and Group "II" to distinguish between them. Group "I", includes 11 glutamine amidotransferases, which are similar in sequence among each other [29] (see Table III). The comparison of the amino acid sequences of this group with the amino acid sequence of the large GOGAT subunit, showed no global statistical significance. Group "II"

includes the amidophosphoribosyltransferases of E. coli and B. subtilis. The comparison of the sequence from the B. subtilis enzyme with the large GOGAT subunit obtained a score of 2.9 SD. This result could suggest the existence of a statistically significant similarity between specific protein segments from the large GOGAT subunit and the amidophosphoribosyltransferase from B. subtilis.

Sequence alignment between the glutamine-binding site of several glutamine aminotranferases and the large GOGAT subunit

An overall similarity of the large GOGAT subunit with the glutamine-binding subunits of glutamine aminotranferases from Group "I" was not found. For this reason a 13-amino acid sequence, which contains a cysteine residue identified to be reactive to glutamine analogs in the anthranilate synthase component II of P. putida and S. marcescens [30, 31], was searched for in the large GOGAT subunit sequence, using the ALIGN program. Table III shows an alignment between these conserved amino acid sequences and a protein segment, located at position 235-247, in the large GOGAT subunit, (cluster B, Fig. 1). In this sequence, 5 out of 13 amino acids are identical to the consensus, and another 4 are conservative substitutions [13]. It is important to remark that although the similarity is not high, the GOGAT sequence contains a cysteine residue that is present in an equivalent position (position 242) to that of the identified reactive cysteine residue in this group of glutamine amidotranferases. However, two other blocks of conserved amino acids, designated regions I and III, which are essential for the

glutamine-dependent activity in these enzymes, [32], were not found at the approximate equivalent position in the GOGAT sequence.

Another type of glutamine amide transfer domain has been reported in the amidophosphoribosyltransferase from E. coli and E. subtilis; this domain is localized at the amino termini of these enzymes [33, 34]. Fig. 3 shows an alignment obtained with the ALIGN program, of the first amino acid residues from the amino termini region of these enzymes and the corresponding amino terminal region in the large GOGAT subunit. The GOGAT sequence shows 26% and 36% similarity to the E. coli and to the B. subtilis amidophosphoribosyltransferase sequences, respectively. It is important to observe that a cysteine residue, required for catalytic activity in both amidophosphoribosyltransferases, is present at the equivalent position in the GOGAT sequence (see Fig. 3).

We have determined the amino acid sequence from the amino terminal region of purified GOGAT from E. coli. The large mature subunit of this enzyme possesses a cysteine residue as the mature amino terminus. There are two possible translation initiation sites preceding this cysteine residue [6]. Experiments from our group, have identified the most probable translation initiation site as the one localized 11 amino acid residues before the aforementioned cysteine residue (unpublished observations) (Fig. 3). The alignment in Fig. 3 also shows that there is sequence similarity between the leader peptide from the amidophosphoribosyltransferase of B. subtilis and the putative leader peptide in the large GOGAT subunit; also, both sequences

have the same size (see Fig 3 and Fig. 1, cluster A).

A specific proteolytic cleavage removes the first 11 amino acid residues from the amidophosphoribosyltransferase of B. subtilis. From such processing, a cysteine residue is exposed as the amino terminus. In the same enzyme from E. coli, after the first methionine is removed, an equivalent cysteine residue is exposed as the amino terminus of the mature enzyme. In both cases, this cysteine residue is the one reactive with the glutamine analog 6-diazo-5-oxo-L-norleucine [33, 34]. In light of these data, it is possible to propose that a proteolytic process similar to the one for the amidophosphoribosyltransferase of B. subtilis, is responsible for generating the mature amino-terminus of the E. coli large GOGAT subunit. It is important to point out that, when the gene that codes for the Bacillus enzyme is introduced in E. coli, the protein product is correctly processed [35]. Therefore, it is tempting to speculate that the protease that processes the amidophosphoribosyltransferase from B. subtilis in E. coli, could be the same one that normally removes the 11 amino acid leader peptide in the GOGAT large subunit. It would be interesting to find out the physiological role of such processing in GOGAT and also if this event occurs in the GOGAT from other organisms.

Considering the sequence similarity, the possible common processing mechanism, and the existence of an amino-terminal cysteine residue, we propose that the amino terminal region of the large GOGAT subunit could be part of the glutamine-binding active site in this enzyme.

The other possible glutamine-binding site previously mentioned (position 236-249, cluster B, Fig. 1) could be involved in the allosteric modulation of this enzyme by glutamine. In this regard, Bower and Zalkin [36] have reported the binding of 0.5 equivalent of 2-oxoglutarate per protomer in the GOGAT from E. coli, in the absence of glutamine. In the presence of this amino acid, one equivalent of 2-oxoglutarate binds per protomer, probably, as suggested by the authors, by the induction of a conformational change of the enzyme [36]. At the present time we are interested in demonstrating if these two putative glutamine-binding sites are functional.

The results presented in this work are part of an initial effort that intends to define, using a systematic comparison analysis of the information already present in protein sequence databanks, possible functional and relevant sites in the GOGAT polypeptides. These results represent the first step in this direction. Further experiments and comparisons should be developed in order to prove if any of these propositions are true. Nevertheless, with these data we know now which polypeptide segments have a greater chance to be involved in some of the substrate and cofactor-binding sites in the GOGAT enzyme from E. coli.

Acknowledgments. We wish to thank the Instituto de Investigaciones Eléctricas (Cuernavaca, México), for allowing us to use their computer facilities. Specially the use of the VAX 11/780 computer. We also wish to thank Dr. Eisenberg and Dr. de Robertis, University of California at Los Angeles, for allowing the use of their computer facilities and for valuable discussions. G. Gosset, E. Merino and F. Recillas are recipients of fellowships from Consejo Nacional de Ciencia y Tecnología, México. This work was supported in part by Consejo Nacional de Ciencia y Tecnología, México, Donativo #PCCBNAL-022584.

References

1. Mäntsälä P, and Zalkin H (1976) *J Biol Chem* 251:3300-3305
2. Tempest DW, Meers JL, Brown EM (1970) *Biochem J* 117:405-407
3. Lea PJ, Mifflin BJ (1975) *Biochem Biophys Res Commun* 64:856-862
4. Roon RJ, Even HL, Larimore F (1974) *J Bacteriol* 118:89-95
5. Mifflin BJ, Lea PJ (1976) *Phytochemistry* 15:873-875
6. Oliver G, Gosset G, Sanchez-Pescador R, Lozoya E, Ku LM, Flores N, Becerril B, Valle F, Bolivar F (1987) *Gene* 60:1-11
7. Trotta PP, Platzter KEB, Haschemeyer RH, Meister A (1974) *Proc Natl Acad Sci USA* 71:4607-4611
8. Miller RE, Stadtman ER (1972) *J Biol Chem* 247:7407-7419
9. Mäntsälä P, Zalkin H (1976) *J Biol Chem* 251:3294-3299
10. Geary LE, Meister AJ (1977) *J Biol Chem* 252:3501-3508
11. Lipman DJ, Pearson WR (1975) *Science* 227:1435-1441
12. Schwartz RM, Dayhoff MO (1979) In: Dayhoff MO (ed) *Atlas of protein sequence and structure, vol 5, suppl 3*. National Biomedical Research Foundation, Washington, DC, pp 353-358
13. Dayhoff MO, Schwartz RM, Orcutt BC (1979) In: Dayhoff MO (ed) *Atlas of protein sequence and structure, vol 5, suppl 3*. National Biomedical Research Foundation, Washington, DC, pp 345-362
14. Chou PY, Fasman GD (1974) *Biochemistry* 13:222-245
15. Chou PY, Fasman GD (1978) *Annu Rev Biochem* 47:251-276
16. Rice DW, Schulz GE, Guest JR (1984) *J Mol Biol* 174:483-496
17. Musick WDL, Rossmann MG (1979) *J Biol Chem* 254:7611-7620
18. Grau UM, Trommer WE, Rossmann MG (1981) *J Mol Biol*

151:289-307

19. Torff HJ, Becker D, Schwarzwald J (1977) In: Sund H (ed) Pyridine Nucleotide Dependent Dehydrogenases, Walter de Gruyter, Berlin, pp 31-42
20. Rossmann MG, Moras D, Olson KW (1974) Nature 250:194-199
21. Wootton JC (1974) Nature 252:542-546
22. Wierenga RK, De Maeyer MCH, Hol WGJ (1985) Biochemistry 24:1346-1357
23. Wierenga RK, De Jong RJ, Kalk KH, Hol WGJ, Drenth J (1979) J Molec Biol 131:55-73
24. Rossmann MG, Argos PJ (1975) J Biol Chem 250:7525-7532
25. Rossmann MG, Liljas A, Branden CI, Banaszak LJ (1975) In: Boyer PD, (ed) The Enzymes, vol 11, Academic Press, New York, pp 61-102
26. Schulz GE (1980) J Mol Biol 138:335-347
27. Valle F, Becerril B, Chen E, Seeburg P, Heyneker H, Bolivar F (1984) Gene 27:193-199
28. Rendina AR, Orme-Johnson WH (1978) Biochemistry 17:5388-5393
29. Nyunoya H, Lusty CJ (1984) J Biol Chem 259:9790-9798
30. Kawamura M, Keim PS, Goto Y, Zalkin H, Heinrikson RL (1978) J Biol Chem 253:4659-4668
31. Tso JY, Hermodson MA, Zalkin H (1980) J Biol Chem 255:1451-1457
32. Amuro N, Paluh JL, Zalkin H (1985) J Biol Chem 260:14844-14849
33. Tso JY, Hermodson MA, Zalkin H (1982) J Biol Chem 257:3532-3536
34. Vollmer SJ, Switzer RL, Hermodson MA, Bower SG, Zalkin H

(1983) J Biol Chem 258:10582-10585

35. Makaroff CA, Zalkin H, Switzer RL, Vollmer SJ (1983) J Biol Chem 258:10586-10593

36. Bower S, Zalkin H (1983) Biochemistry 22:1613-1620

Figure 1. Detected similarities between GOGAT and protein sequences compiled in the National Biomedical Research Foundation Database.

The linear polypeptide maps of the two GOGAT subunits are represented by the two large continuous lines. The large and small subunits are composed by 1514 and 471 amino acid residues respectively [6]. Related protein segments are represented by a line over the GOGAT region with which it shares similarity. There are five clusters of protein segments that share similarity with specific regions within GOGAT. Clusters A and B were found using the ALIGN program. Cluster E and part of cluster C were found using the FASTP program. Clusters D and part of C were found using the RELATE program. Cluster A includes the amino-terminal segments from the glutamine phosphoribosylpyrophosphate amidotransferase from E. coli and B. subtilis. Cluster B includes 11 homologous segments from different glutamine amidotransferases. Clusters C, D and E include segments from proteins that bind a dinucleotide cofactor. Names of proteins whose segments are present in clusters A to E:

CLUSTER A

- 1- Glutamine phosphoribosylpyrophosphate amidotransferase E. coli
- 2- Glutamine phosphoribosylpyrophosphate amidotransferase B. subtilis

CLUSTER B

- 1- Anthranilate synthase, component II Escherichia coli
- 2- Anthranilate synthase, component II Salmonella typhimurium
- 3- Anthranilate synthase, component II Serratia marcescens
- 4- Anthranilate synthase, component II Pseudomonas putida

- | | |
|---|------------------------------------|
| 5- Anthranilate synthase, component II | <u>Acinetobacter calcoaceticus</u> |
| 6- Anthranilate synthase, component II | <u>Saccharomyces cerevisiae</u> |
| 7- Anthranilate synthase, component II | <u>Shigella dysenteriae</u> |
| 8- Anthranilate synthase, component II | <u>Neurospora crassa</u> |
| 9- Carbamoyl-phosphate synthase, small chain | <u>Saccharomyces cerevisiae</u> |
| 10- Carbamoyl-phosphate synthase, small chain | <u>Escherichia coli</u> |
| 11- p-Aminobenzoate synthase, component II | <u>Escherichia coli</u> |

CLUSTER C

- | | |
|---|--------------------------------|
| 1- P-Hydroxybenzoate hydroxylase | <u>Pseudomonas fluorescens</u> |
| 2- Lipoamide dehydrogenase | <u>Escherichia coli</u> |
| 3- Mercuric reductase (transposon Tn 501) | <u>Pseudomonas aeruginosa</u> |
| 4- Lactate dehydrogenase M chain | <u>Gallus gallus</u> |
| 5- Lactate dehydrogenase H chain | <u>Sus scrofa domestica</u> |
| 6- Lactate dehydrogenase X chain | <u>Mus musculus</u> |

CLUSTER D

- | | |
|---|--------------------------------|
| 1- P-Hydroxybenzoate hydroxylase | <u>Pseudomonas fluorescens</u> |
| 2- Lipoamide dehydrogenase | <u>Escherichia coli</u> |
| 2- Mercuric reductase (transposon Tn 501) | <u>Pseudomonas aeruginosa</u> |
| 4- Lactate dehydrogenase M chain | <u>Gallus gallus</u> |
| 5- Lactate dehydrogenase H chain | <u>Sus scrofa domestica</u> |
| 6- Lactate dehydrogenase X chain | <u>Mus musculus</u> |

CLUSTER E

- | | |
|---|-------------------------------|
| 1- Lipoamide dehydrogenase | <u>Escherichia coli</u> |
| 2- Mercuric reductase (transposon Tn 501) | <u>Pseudomonas aeruginosa</u> |

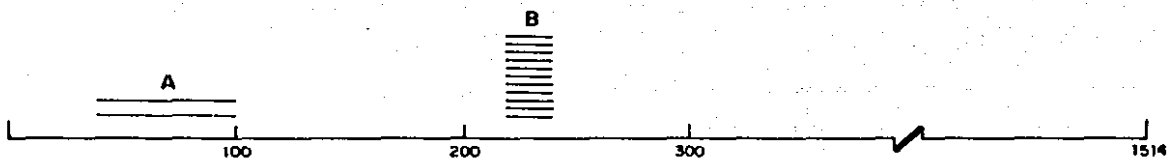
Figure 2. Secondary structure prediction profile of a protein segment (140-430) from the small GOGAT subunit.

This figure shows the secondary structure profile of the protein region (140-430) from the small GOGAT subunit, that includes the two segments ("c" and "d") shown in Table II. The Chow and Fasman method [14, 15] for the prediction of secondary structure was used with a window size of 6 residues. The average scores are plotted at the position corresponding to the middle of each segment. The Y axis represents the range of conformational parameters for prediction of secondary structure. Continuous and broken lines represent probability for beta-sheet and alpha-helix structure formation respectively. The X axis shows the amino acid sequence position in the GOGAT small subunit. Line I indicates the predicted secondary structure elements. Hatched areas represent beta-turn regions. The two boxes shown in line II, represent the two GOGAT protein segments ("c" and "d") that share similarity to the dinucleotide binding domains present in several proteins (see Table I).

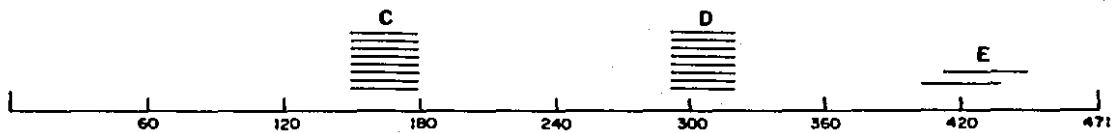
Figure 3. Amino acid sequence alignment of the amino-terminal regions from glutamine phosphoribosylpyrophosphate amidotransferase of E. coli and B. subtilis and the large GOGAT subunit.

Identical and equivalent amino acids are boxed; conservative replacements: E-D, I-L-V-F-M, H-Q [13]. The standard single letter code is used. Asterisks indicate the translation initiation site in the glutamine phosphoribosylpyrophosphate amidotransferase enzymes, and the most probable initiation site

in GOGAT. The vertical arrow indicates the site where proteolytical processing occurs in these three enzymes. The identified active cysteine residue in the glutamine phosphoribosylpyrophosphate amidotransferase from E. coli and B. subtilis is indicated by a small dot.



Large GOGAT
Subunit



Small GOGAT
Subunit

Fig. 1

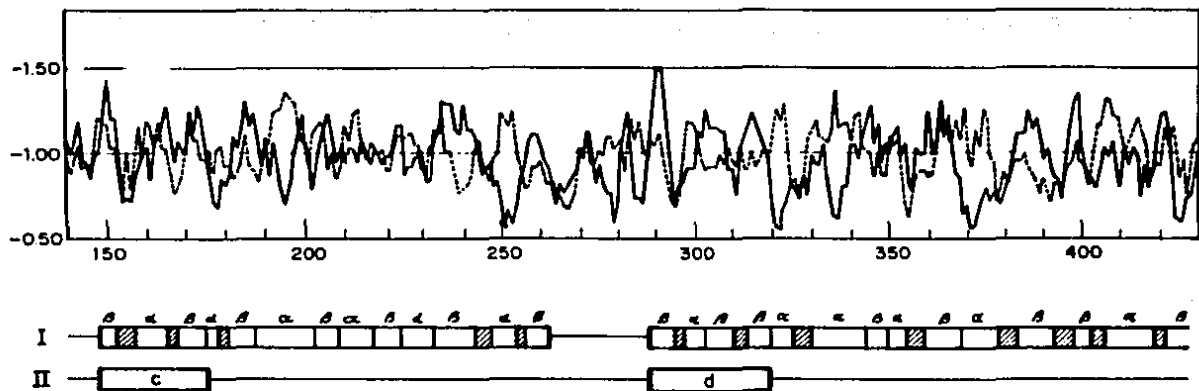
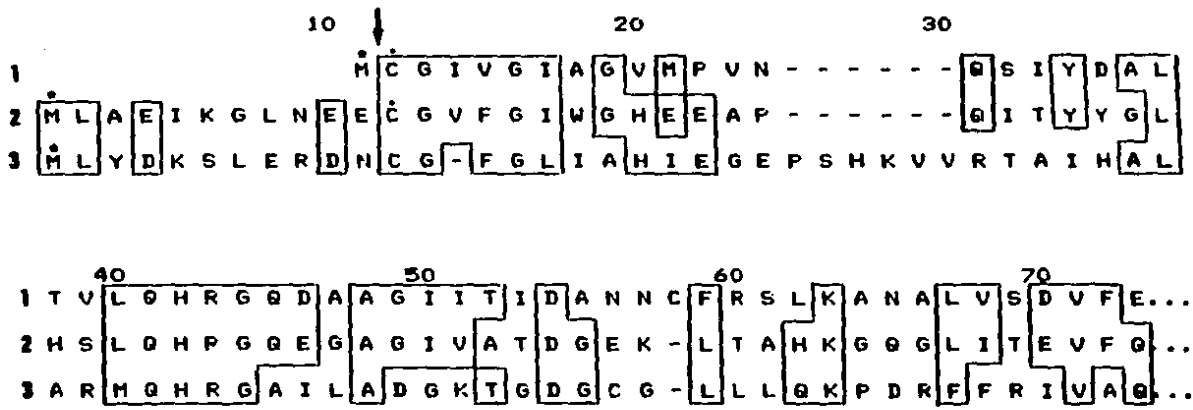


Fig. 2



1. Glutamine phosphoribosylpyrophosphate amidotransferase, *E. coli*.
2. Glutamine phosphoribosylpyrophosphate amidotransferase, *B. subtilis*.
3. Glutamate synthase, *E. coli*.

Fig. 3

TABLE I

Statistical comparison analysis between the small GOGAT subunit and six enzymes that bind a dinucleotide cofactor.

Score*	Top scores considered	Enzyme	Organism
5	20	P-Hydroxybenzoate hydroxylase	<u>Pseudomonas fluorescens</u>
4.3	50	Lipoamide dehydrogenase	<u>Escherichia coli</u>
3.2	35	Mercuric reductase	<u>Pseudomonas aeruginosa</u>
3.9	40	Lactate dehydrogenase M chain	<u>Gallus gallus</u>
3.2	35	Lactate dehydrogenase H chain	<u>Sus scrofa domestica</u>
3.2	40	Lactate dehydrogenase X chain	<u>Mus musculus</u>

* Scores were obtained using the RELATE program and are expressed in standard deviation units. Scores higher than 3 SD have been defined as statistically significant [13].

TABLE II

Sequence similarity between segments from the dinucleotide-binding-folds of the six enzymes presented in Table I and two protein segments from the small GOGAT subunit.

c.	K V A I I G A G P A G L A C A D V L T R N G V K A - V C F D R H
d.	R V V V L G G G D T A M D C V R T S V R Q G A N D - V T C A Y R
1.	Q V V V L G A G P A G Y S A A F R C A D L G L E T - V I V E R Y
2.	Q V A I I G A G P S G L L L G Q L L H K A G I D N - V I L E R Q
3.	Q V A V I G S G G A A M A A A L K A V E Q G A Q V - T L I E R G
4.	R L L V M G G G I I G L E M G T V Y H A L G S Q I - D V V E M F
5.	R L A V I G S S V V A L E L A Q A F A R L G S K V T V L A R N
6.	K I T V V G V G B V G M A C A I S I L L K G L A D E L A L V D
7.	K I T V V G V G Q V G M A C A I S I L G K S L T D E L A L V D
8.	K I S V V G V G A V G M A C A I S I L M K D L A D E L T L V D
	H H G G G H H H H H

Small GOGAT subunit:	Segment position
c. Segment "c"	(148-178)
d. Segment "d"	(289-319)

Dinucleotide-binding enzymes:	Segment position	Cofactor
1 Lipoamide dehydrogenase (<u><i>E. coli</i></u>)	(6-36)	FAD
2 P-hydroxybenzoate hydroxylase (<u><i>P. fluorescens</i></u>)	(4-34)	FAD
3 Mercuric reductase (<u><i>P. aeruginosa</i></u>)	(100-130)	FAD
4 Lipoamide dehydrogenase (<u><i>E. coli</i></u>)	(176-206)	NAD
5 Mercuric reductase (<u><i>P. aeruginosa</i></u>)	(272-302)	NADP
6 Lactate dehydrogenase (<u><i>Mus musculus</i></u>)	(20-50)	NAD
7 Lactate dehydrogenase (<u><i>Sus scrofa domestica</i></u>)	(22-52)	NAD
8 Lactate dehydrogenase (<u><i>Gallus gallus</i></u>)	(21-51)	NAD

Table II. The local beta-alpha-beta units of several FAD-NADP-binding-folds from the enzymes present in Table I are aligned to show the similarities between these segments and two sequences located in the small GOGAT subunit: segments "c" (148-178) and "d" (289-319). Conserved amino acid residues are indicated below the expected position on the beta-alpha-beta units. "H"= conserved hydrophobic amino acid residues defined by Wierenga et al. [23] as important component residues in the beta-alpha-beta unit of a dinucleotide-binding domain. "G"= conserved glycine residues in the beta-alpha-beta unit of the dinucleotide-binding domain.

TABLE III

Alignment of a 13 amino acid protein segment present in the glutamine binding subunit of certain glutamine amidotransferases.®

Sequence	Position	Protein #
L P I I G I C L G H Q A I	78-90	<u>E. coli</u> ASII
L P I I G I C L G H Q A I	77-89	<u>S. typhimurium</u> ASII
L P I I G I C L G H Q A I	77-89	<u>S. dysenteriae</u> ASII
L P I I G I C L G H Q A I	77-89	<u>S. marcescens</u> ASII
L P I L G V C L G H Q S I	73-85	<u>P. putida</u> ASII
I P V F G I C M G Q Q C M	86-98	Yeast ASII
I P I F G V C M G Q Q C I	98-110	<u>N. crassa</u> ASII
I P L L G V C L G H Q A I	73-85	<u>Acinetobacter calcoaceticus</u> ASII
I P V F G I C L G H Q L L	263-275	<u>E. coli</u> CPS
I P I F G I C L G H Q L L	258-270	Yeast CPS
L P I L G V C L G H Q A M	73-85	<u>E. coli</u> PABSII
L L P I I G I C L G H Q A I I F V		Conserved amino acids
A S G M A I C L F H Q R F	236-249	<u>E. coli</u> GOGAT large subunit

® Equivalent amino acids were defined from Dayhoff's mutational data matrix, conservative replacements: E-D, I-L-V-F-M, H-Q [13].

ASII, anthranilate synthase Component II; CPS, carbamyl phosphate synthetase small subunit; PABSII, p-aminobenzoate synthase Component II.

* Cysteine residue identified as part of the glutamine-active site.

RESULTADOS ADICIONALES Y DISCUSION

Organización transcripcional de los genes *gltB* y *gltD*

Los resultados de los experimentos de extensión de primero, utilizando un oligonucleótido que hibridaba próximo a la región flanqueadora 5' de *gltB*, indican que en las condiciones de crecimiento usadas, para la extracción del RNA (medio NN suplementado con amonio 15 mM como fuente de nitrógeno), la transcripción proviene del promotor localizado en la posición 146-174 (ver Fig. 5 y Fig. 1, Oliver et al., 1987). Utilizando un primero que hibridaba en la región 5' de *gltD* y bajo las mismas condiciones experimentales que en caso anterior, no se logró detectar inicio de la transcripción proveniente de la región flanqueadora 5' para el gene estructural de la subunidad pequeña (dato no publicado).

Por otro lado, los experimentos de hibridación RNA-DNA, revelan la existencia de un transcrito de aproximadamente 7.3 kb, el cual contiene información proveniente de *gltB* y de *gltD* (Fig. 4, Oliver et al., 1987). Este transcrito ha sido detectado en condiciones de crecimiento en medio mínimo con amonio 15 mM o 0.5 mM como fuente de nitrógeno. Estos datos indican que los genes estructurales para GOGAT se encuentran organizados formando un operón o unidad transcripcional. Es importante señalar que el tamaño del transcrito resulta ser mayor del esperado (aprox.

6 kb), si solo codificara para gltB y gltD. Recientemente se ha reportado que este operón contiene un tercer gene (gltF), localizado en la región 3' de la unidad transcripcional (Castaño et al., 1988). Este dato indica que el RNA mensajero para GOGAT es tricistrónico.

Las técnicas experimentales usadas no permitirían identificar claramente a un transcrito único para gltD, ya que frecuentemente se detectaba hibridación con los RNAs ribosomales. El RNAr 16S tiene un tamaño similar al del posible transcrito para gltD, por lo tanto los experimentos de hibridación RNA-DNA difícilmente podrían usarse como evidencia para establecer la existencia de un transcrito único para el gene estructural de la subunidad pequeña.

Los datos presentados indican que la región flanqueadora 5' de gltB es la región de promoción de este operón en condiciones de crecimiento en un medio mínimo suplementado con alto o bajo amonio. Sin embargo, esto no descarta la posibilidad de que gltD pudiera contar también con transcripción específica bajo ciertas condiciones. También es posible la existencia de un transcrito específico solo para gltB, el cual se generaría por un evento de terminación en el extremo 3' del gene estructural.

Es importante considerar que la enzima GOGAT está formada por cantidades equimolares de las dos subunidades. Este dato se relaciona con el hecho de que existe una regulación coordinada de la transcripción de los dos genes, al estar estos integrados en un operón. La determinación de la secuencia de DNA para gltB y gltD indica que se encuentran separados únicamente por 12

pares de bases. Por lo tanto el codón de terminación para gltB se encuentra muy cercano al codón de iniciación para gltD (Fig. 2, Oliver et al., 1987). Esto sugiere que los dos genes pudieran estar acoplados traduccionalmente, favoreciéndose de esta manera una expresión coordinada que permita la síntesis de cantidades similares para las dos subunidades (Oppenheim y Yanofsky, 1980).

Por otro lado, la separación de 12 pares de bases entre las dos subunidades permitiría que una mutación supresora del codón de terminación de gltB, ocasione la fusión de las dos subunidades. Es importante señalar que la GOGAT de N. crassa es un solo polipéptido de aproximadamente 220 kDa. La fusión de las dos subunidades de la GOGAT de E. coli resultaría en una proteína de aproximadamente 218 kDa. Este dato es interesante ya que podría explicar el mecanismo por el cual se originó la GOGAT de un organismo como N. crassa. Esta suposición podrá ser apoyada cuando se conozca la secuencia del gene para la enzima de N. crassa.

La localización del promotor funcional para el operón, bajo condiciones de crecimiento en alto amonio, permite relacionarlo con las secuencias regulatorias previamente propuestas. En la Fig. 2 del artículo de Oliver et al. (1987), se pueden observar la posición del promotor y de algunas de las regiones que pudieran estar involucradas en regulación, las cuales se describen en el mismo artículo.

Se detectaron por criterio de consenso dos posibles sitios de unión para el complejo AMPC-CRP, centrados en las posiciones 152-153 y 239-240. Uno de estos dos posibles sitios sobrelapa la región -35 del promotor funcional, este dato puede ser relacionado con un reporte que indica la reducción de la actividad de GOGAT, en células con un alto nivel del complejo AMPC-CRP (Prusiner et al., 1972). Se identificó también una secuencia palindrómica (posición 197-215, Fig. 2, Oliver et al. 1987). Este tipo de secuencia es común en operadores o en regiones reconocidas por proteínas que se unen al DNA. La posición del palíndrome con respecto al promotor es poco frecuente, ya que su eje de simetría se encuentra a 29 pares de bases del inicio de la transcripción y por lo tanto formaría parte del transcrito. Experimentos recientes en nuestro laboratorio indican que el inicio de la traducción más probable para la subunidad grande es el que se encuentra en la posición 394 (Fig. 2, Oliver et al., 1987). Esto indica que el líder del transcrito para GOGAT tiene una longitud de 216 bases.

Con el fin de saber si las secuencias identificadas en GOGAT presentaban similitud con otras que hayan sido caracterizadas, se procedió a comparar la región de regulación de GOGAT con las de otros genes. En la Fig. 6 de Oliver et al. (1987) se muestra un alineamiento entre una parte del líder del transcrito para GOGAT y el líder del operón rrnB de E. coli. Es importante señalar que la similitud incluye a la secuencia "Box A", la cual se encuentra implicada en eventos de antiterminación en operones ribosomales bacterianos y genes de bacteriófagos (Olson et al., 1984).

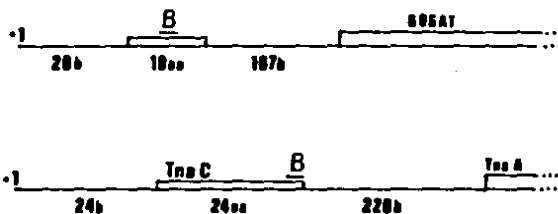
La presencia de una secuencia con un posible papel de antiterminador en la región líder del operón de GOGAT, pudiera explicarse si se considera que puede cumplir con alguna de las siguientes funciones:

1.- En los operones ribosomales, la antiterminación permite la transcripción de moléculas de RNA ribosomal las cuales son muy estructuradas. Esta estructuración pudiera interferir con el progreso de la RNA polimerasa, siendo necesario un mecanismo que permita la transcripción a través de terminadores fortuitos. Puesto que el RNA mensajero de GOGAT es excepcionalmente grande (aproximadamente 7.3 Kb), es probable que la antiterminación sea un mecanismo que pudiera asegurar la transcripción completa del operón.

2.- En el extremo 3' de gltB (posición 4826-4841, Fig. 2, Oliver et al., 1987) se identificó una secuencia palindrómica que presenta una estructura similar a la del terminador dependiente de rho tr1 del fago lambda. Una terminación diferencial en ese sitio, permitiría la síntesis de transcritos completos para el operón, o transcritos solo para gltB. Estos eventos estarían determinados por antiterminación regulada, de una manera similar a la existente en el fago lambda.

Las dos posibles funciones mencionadas arriba, no son excluyentes. Sin embargo, es importante aclarar que hasta el momento no se cuenta con datos experimentales que apoyen algún evento de antiterminación en la transcripción del operón.

Uno de los operones que posee en su región líder una secuencia de tipo "Box A" es el operón tna, el cual contiene los genes que codifican para la enzima triptofanasa y para una permeasa con baja afinidad al triptofano. Se comparó la región de regulación de tna, con la región equivalente en GOGAT. A continuación se muestra la figura I en donde se representa la organización de estas regiones:



GOGAT

+1 AUCCGCGGAAAGCUUUCUGGAUGAGCAGCCUGGCUCAUCAUAAUUUAUGCAGUAAUUGAGAU...g1tB
 METSerSerLeuLeuIleIlePheMETG1n .

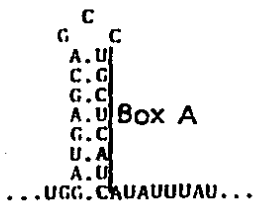


Fig. I

En la parte superior de la figura se comparan las regiones 5' de los transcritos identificados para los operones de GOGAT y tna. En la región líder de tna se ha identificado una secuencia que es traducida (tnaC) y que sobrelapa con la secuencia "Box A", indicada con una "B". El tamaño del péptido líder codificado por tnaC es de 24 aminoácidos. Se trató de identificar una organización similar en la región equivalente del transcrito para GOGAT. En la posición 198-228 (Fig. 2, Oliver et al., 1987), se localizó una fase abierta de lectura que podría codificar para un péptido de 10 aminoácidos. En la parte inferior de la figura I se muestra la secuencia de los primeros 60 nucleótidos del transcrito para GOGAT producido bajo condiciones de crecimiento con amonio 15 mM. Se puede observar que el ATG de la fase abierta de lectura previamente mencionada, está precedido a una distancia adecuada (Stormo et al., 1982) por una secuencia con características similares a un sitio de unión a ribosoma, señalada por puntos sobre los residuos que son complementarios al RNA ribosomal 16S. Es importante notar que el posible péptido tendría residuos repetidos (Ser-Ser-Leu-Leu-Ile-Ile), en una disposición similar a la encontrada en los péptidos regulatorios de algunos atenuadores (Bauer et al., 1983). Los aminoácidos serina, leucina e isoleucina, tienen en común que requieren durante el último paso de su síntesis, de un grupo amino proveniente del glutamato. La secuencia palindrómica se encuentra señalada por flechas y además se muestra también la estructura secundaria que podría adquirir; la posible secuencia "Box A" formaría parte del tallo. Se debe mencionar que después

de la estructura de tallo y asa, existe una alta proporción de residuos de uridina, mostrando cierta similitud con un terminador (Adhya y Gottesman, 1978). Es importante observar que los tres posibles elementos regulatorios: la secuencia palindrómica, "Box A" y la fase abierta de lectura, se encuentran sobrelapados en la misma región del transcrito.

La similitud en organización entre las regiones regulatorias de tna y GOGAT, sugiere que pudiera existir un mecanismo de regulación análogo. Lo que se sabe respecto de la regulación en tna es que concentraciones altas de triptofano en el medio inducen al gene estructural para la triptofanasa (tnaA). Se ha identificado un terminador dependiente de rho inmediatamente después de la región que codifica para tnaC. En condiciones de inducción se propone que el triptofano se uniría a una proteína hipotética, la cual va a interaccionar con "Box A" y permitir la transcripción a través del terminador dependiente de rho. (Stewart y Yanofsky, 1985).

Por el momento resulta difícil proponer un esquema que integre los elementos antes descritos, para explicar la regulación de los genes estructurales para GOGAT. Es necesario antes realizar experimentos que demuestren la funcionalidad de tales elementos. En la sección de PERSPECTIVAS de esta tesis, se presentan estrategias de estudio, algunas de las cuales ya se encuentran en progreso.

Es interesante resaltar la posibilidad de que la transcripción de los genes para GOGAT pudiera estar regulada no directamente por el glutamato, sino por aminoácidos que lo requieren como

precursor. Esta propuesta concuerda con el hecho de que el glutamato puede considerarse como uno de los primeros elementos de una vía de biosíntesis, la cual culmina con la formación de varios compuestos nitrogenados en la célula. Por lo tanto es posible pensar que la regulación de una enzima que sintetiza a uno de los primeros compuestos de la vía metabólica, esté regulada por alguno de los productos finales de la misma.

Estudio de las dos subunidades que conforman a GOGAT

En la sección de INTRODUCCION, se incluyen los datos y conclusiones más importantes, obtenidos de estudios de la cinética de las reacciones catalizadas por GOGAT. Algunas de las actividades identificadas para cada una de las dos subunidades de GOGAT, así como los sustratos y cofactores utilizados por las mismas, son compartidos por otras enzimas. Se conoce la secuencia polipeptídica de algunas de éstas y se ha logrado, en ciertos casos, determinar su estructura tridimensional por análisis de difracción de rayos X. Esta información, junto con la obtenida de estudios cinéticos y de unión de ligandos, ha permitido establecer que secuencias o residuos de aminoácidos intervienen en la catálisis o la unión de sustratos, cofactores, grupos prostéticos o moléculas moduladoras de la actividad en esas enzimas.

La mayoría de las secuencias de aminoácidos determinadas para varias proteínas conocidas, junto con alguna de la información mencionada arriba, se encuentran depositadas en bancos de datos. La posibilidad de contar con la secuencia de aminoácidos deducida para las dos subunidades de GOGAT, permitió realizar un estudio basado en la detección de similitudes con secuencias de proteínas compiladas en bancos de datos. La detección de

similitud entre regiones en GOGAT y regiones de otras proteínas, las cuales se ha demostrado realizan una función específica, permitiría asignar a la secuencia en GOGAT una función similar.

El primer paso en este estudio consistió en terminar de secuenciar el gene estructural para la subunidad pequeña de GOGAT. Se secuenciaron aproximadamente 700 bases de la región 3' de gltD. La determinación de la secuencia completa de los genes estructurales para las dos subunidades de GOGAT, permitió deducir la secuencia de aminoácidos de éstas, para compararlas con las secuencias de otras proteínas tal y como se describe en las secciones de "Materials and methods" y "Results" en la publicación "Amino acid sequence analysis of the glutamate synthase enzyme from Escherichia coli K-12" de Gosset et al., (1988). A continuación se presenta una discusión de los resultados obtenidos en este estudio, considerando primero a la subunidad pequeña y después a la subunidad grande.

Los resultados de los análisis usando los programas FASTP, RELATE y ALIGN, indican que la subunidad pequeña de GOGAT posee tres regiones de similitud con seis enzimas que participan en reacciones de oxidorreducción (Fig. 1, grupos C, D y E; Tabla I, Gosset et al., 1988). Dos regiones en la subunidad pequeña de GOGAT (segmentos "c" y "d", Tabla II, Gosset et al., 1988) muestran similitud hacia dominios con la función de unir dinucleótido en las seis enzimas antes mencionadas. Un examen detallado de la secuencia de los segmentos "c" y "d", reveló que ambos poseen residuos específicos, conservados en los sitios de unión a dinucleótido de las seis enzimas. A tres de estas se

les ha determinado su estructura tridimensional mediante estudios de difracción de rayos X (P-hidroxibenzoato hidroxilasa, lactato deshidrogenasa cadena H y lactato deshidrogenasa cadena X; Tabla II, Gosset et al., 1988).

A partir de estudios de la estructura tridimensional de varias enzimas que utilizan dinucleótidos como cofactores, se han definido los elementos estructurales mínimos que conforman a un sitio de unión a dinucleótido (Wierenga et al., 1985). Algunos de estos elementos pueden ser identificados a nivel de la estructura secundaria del dominio y son: a) la presencia de la llamada unidad beta-alfa-beta, donde se encuentran los residuos específicos más conservados y b) una alternancia de estructuras alfa y beta con pocas regiones no estructuradas en un segmento que abarca aproximadamente 100-120 residuos de aminoácidos (Rossmann et al., 1974; Wootton, 1974). La predicción de la estructura secundaria de la subunidad pequeña de GOGAT, permitió la búsqueda de dichos elementos (Fig. 2, Gosset et al., 1988). Los resultados de este análisis indican la presencia de los elementos antes mencionados, aún en las regiones donde no se detectó similitud a nivel de secuencia primaria entre los segmentos "c" y "d" y los sitios de unión a dinucleótido de las seis enzimas (Tabla II, Gosset et al., 1988).

Todos estos datos sugieren la existencia de dos sitios de unión a dinucleótido en la subunidad pequeña de GOGAT. Por el tipo de reacción que cataliza esta subunidad (ver INTRODUCCION), se esperaba identificar un solo sitio de unión a

dinucleótido. Sin embargo, la presencia de dos sitios puede deberse a que ambos son requeridos para cumplir alguna de las funciones discutidas a continuación:

1.- Revisando la literatura, se encontró que la presencia de dos sitios de unión en una enzima no es un caso poco común. Dos de las proteínas presentadas en la Tabla 11 poseen dos sitios de unión a dinucleótido (lipoamido deshidrogenasa y reductasa mercúrica), uno de los cuales es específico para NAD(P) y el otro para FAD. Sin embargo, no existe evidencia que indique la presencia de FAD en la subunidad pequeña de GOGAT. Por lo tanto resulta difícil proponer esa función para uno de los dos posibles sitios encontrados.

2.- Las enzimas glutamato deshidrogenasa de bovino y de N. crassa, que solo unen NADPH, presentan también dos sitios de unión al dinucleótido, uno de los cuales se acerca más al consenso general para este tipo de dominios. En este caso, ambos dominios son requeridos para la unión del cofactor NADPH. Se podría pensar en un mecanismo similar para GOGAT, en que ambos sitios sean necesarios; aún cuando uno de ellos (segmento "d") se encuentre menos conservado.

3.- Estudios realizados por Rendina y Orme-Johnson (1978), demuestran la activación de la actividad de GOGAT por NADP⁺ cuando los niveles de alfa-cetoglutarato son bajos. La activación es consecuencia de una disminución de la Km por el alfa-cetoglutarato, y posiblemente sea debida a la interacción del NADP⁺ en un sitio distinto al de unión para NADPH. Ese

trabajo sugiere la existencia de dos sitios de unión para dinucleótido en GOGAT, uno de los cuales tendría un papel modulador de la actividad enzimática.

Aunque las primeras dos funciones para dos sitios de unión son las que se han demostrado, pudieran existir otras aún no descritas para un sitio adicional. Otro punto a considerar es que posiblemente una proteína ancestral, de la cual se originó la subunidad pequeña, requería de dos sitios de unión a dinucleótido. Ahora que esta enzima forma parte de un sistema del cual es una subunidad, posiblemente ya no requiere dos sitios y por esta razón la secuencia de uno de ellos ha divergido y tal vez ya no es funcional.

La semejanza tanto en actividad como en peso molecular entre la subunidad pequeña de GOGAT y la GDH de E. coli ha sido la base de suposiciones respecto a una posible relación evolutiva entre ellas. Mantsala y Zalkin (1976b) utilizando técnicas de fijación de complemento y de doble precipitación por anticuerpo, no lograron detectar relación entre las dos proteínas. La comparación de las estructuras primarias de ambas enzimas con el programa RELATE obtuvo un valor de 2.23 SD el cual no permite apoyar un origen común. Es importante mencionar que la comparación de la secuencia de la GDH de E. coli con la de N. crassa, revela que la primera también posee dos sitios de unión a dinucleótido (datos no publicados). Esto puede explicar porque aunque el valor obtenido entre GDH y GOGAT no es estadísticamente significativo, sí es más alto que la mayoría de las comparaciones entre GOGAT y cualquier otra proteína del

banco. Sin embargo, no es suficiente para que la similitud que pudiera existir a nivel estructural, debido a la presencia en los dos casos de dos sitios de unión a dinucleótido, sea detectada por técnicas inmunológicas.

Se ha propuesto un origen común para el dominio de pegado a mononucleótico presente en varias proteínas. Una duplicación de este dominio, pudo dar origen al motivo estructural de pegado a dinucleótido (Rossmann y Argos, 1975). Igualmente, se ha propuesto que ciertas proteínas se originaron por la fusión de módulos estructurales con funciones específicas (Rossmann et al., 1975). Los datos presentados sugieren a la fusión o duplicación de regiones específicas de DNA, como la causa de los dos sitios de unión a dinucleótido en la subunidad pequeña de GOGAT. Otro origen puede ser la divergencia a partir de una proteína que ya tuviera dos sitios de unión, posiblemente una proteína ancestral relacionada a las que se presentan en la Tabla I de Gosset et al., 1988. Esta suposición es apoyada por el hecho de que los pesos moleculares de estas seis enzimas y el de la subunidad pequeña de GOGAT, se encuentra alrededor de 51 kDa. Además, la similitud entre los segmentos "c" y "d", no es significativamente mayor que la similitud entre cualquiera de estas dos regiones y alguno de los dominios beta-alfa-beta presentados en la Tabla II de Gosset et al., 1988.

El análisis utilizando los programas ALIGN y RELATE, entre la subunidad grande de GOGAT y las secuencias obtenidas por la búsqueda con FASTP, no detectó una similitud estadísticamente significativa a nivel global entre ellas.

Se decidió comparar, utilizando el programa RELATE, la secuencia de la subunidad grande de GOGAT con las secuencias de dos grupos de glutamino amidotransferasas (grupo I y grupo II, ver sección "Results and discussion", Gosset et al., 1988) las cuales no aparecieron en la búsqueda con FASTP. Los resultados de estas comparaciones indican que a nivel global solo existe una similitud importante hacia un miembro del grupo II, la amidofosforribosiltransferasa de *B. subtilis*. Ya que no existe una similitud global entre GOGAT y la mayoría de las glutamino amidotransferasas, se decidió hacer la búsqueda de regiones específicas entre estas enzimas, que se ha demostrado forman parte del sitio de unión a glutamina. Esta búsqueda se efectuó con el programa ALIGN.

La subunidad que une glutamina en las enzimas del grupo I, posee una región de 13 residuos de aminoácidos la cual contiene un residuo de cisteína que se ha demostrado reacciona con análogos de glutamina (Kawamura et al., 1978; Tso et al., 1980). Esta región se ha identificado como el sitio de unión para glutamina y forma parte del sitio activo de la enzima. En la subunidad grande de GOGAT se localizó una secuencia similar al consenso derivado de 11 sitios de unión a glutamina, de igual número de enzimas del grupo I (posición 235-247, Oliver et al. 1987 y Tabla III Gosset et al., 1988). La secuencia en GOGAT

posee 5 de los 13 aminoácidos del consenso, y además, otros 4 residuos son substituciones conservativas (Dayhoff et al., 1979). Es importante señalar que uno de los residuos identicos al consenso es una cisteina, la cual se encuentra en una posición equivalente con respecto a los otros 12 residuos, a la ocupada por la cisteina reactiva identificada experimentalmente en dos de las enzimas de la Tabla III de Gosset et al., (1988). Sin embargo, otras dos regiones que se sabe también conforman al sitio de unión a glutamina en este grupo de glutamino amidotransferasas (Amuro et al., 1985), no se localizaron en la secuencia de GOGAT.

Las enzimas del grupo II, que están constituidas por un polipéptido único, poseen otro tipo de dominio para la unión de glutamina, el cual se encuentra en la región amino terminal (Tso et al., 1982; Vollmer et al., 1983). En la Fig. 3 de Gosset et al., (1988) se muestra un alineamiento entre las regiones amino terminal de las dos enzimas del grupo II y la región equivalente en la subunidad grande de GOGAT. Considerando las substituciones conservativas, la secuencia de GOGAT muestra un 26% y un 36% de similitud hacia las secuencias de las amidofosforribosiltransferasas de E. coli y de B. subtilis respectivamente.

Se ha determinado la secuencia de aminoácidos correspondiente a la región amino terminal de la subunidad grande de GOGAT, identificándose a un residuo de cisteina como el extremo amino terminal de la subunidad (Oliver et al., 1987). Algunos experimentos en nuestro laboratorio indican que el inicio de la

traducción para esta subunidad se encuentra localizado a 11 residuos de aminoácidos antes de la cisteína amino terminal. En la Fig. 3 de Gosset et al., (1988) se puede observar que existe similitud entre los péptidos líderes de la amidofosforribosiltransferasa de B. subtilis y la subunidad grande de GOGAT, además ambos son del mismo tamaño.

Tanto la similitud en la secuencia como la presencia de un residuo de cisteína en la posición amino terminal entre las tres proteínas de la Fig. 3 de Gosset et al., (1988), sugieren que esta región en la subunidad grande de GOGAT pudiera formar parte del sitio catalítico de unión a glutamina.

Se ha reportado la unión de medio equivalente de alfa-cetoglutarato por dímero de GOGAT en ausencia de glutamina. En presencia de este aminoácido, se observa ahora la unión de un equivalente de alfa-cetoglutarato por dímero. Este efecto ha sido atribuido a un cambio conformacional en la enzima, provocado por la unión de la glutamina. Una hipótesis atractiva sería que la secuencia de 13 aminoácidos similar al sitio de unión a glutamina de las enzimas del grupo I, de ser funcional, pudiera tener un papel modulador de la actividad de GOGAT.

Procesamiento de la subunidad grande de GOGAT

Al determinar la secuencia de aminoácidos correspondiente a la región amino terminal de la subunidad grande de GOGAT, se encontró que poseía un residuo de cisteína en su extremo amino terminal (Fig. 2 de Oliver et al., 1987). Puesto que el inicio de traducción más probable para esta subunidad se encuentra a 11 residuos de aminoácido de esa cisteína, es necesario algún tipo de procesamiento para generar el extremo amino terminal de la subunidad madura.

Integrando los datos obtenidos experimentalmente, así como el análisis de la secuencia de aminoácidos, es posible proponer un mecanismo que explique la presencia del residuo de cisteína como extremo amino terminal la subunidad grande de GOGAT.

La similitud entre las regiones amino terminales de la subunidad grande de GOGAT y la amidofosforribosiltransferasa de B. subtilis, incluye al péptido líder de esta última y a un péptido similar en GOGAT. Un corte proteolítico remueve los primeros 11 aminoácidos de la enzima de B. subtilis, dejando a un residuo de cisteína como el extremo amino terminal de la enzima. Tal cisteína reacciona con análogos de glutamina, y por lo tanto se le ha propuesto como integrante del sitio activo de la enzima.

Cuando el gene que codifica para la amidofosforribosiltransferasa de B. subtilis es introducido en E. coli, se encuentra que el producto es procesado correctamente. Considerando que la amidofosforribosiltransferasa de E. coli ya no posee la secuencia correspondiente al péptido líder (ver Fig. 3 de Gosset et al., 1988), es inesperado encontrar en E. coli la actividad que remueve al péptido líder de la enzima de B. subtilis. Esto se explicaría si tal actividad es requerida para procesar a péptidos similares que se encuentran en alguna(s) proteína(s) de E. coli. Es posible pensar que la proteasa que procesa al péptido líder de la subunidad grande de GOGAT pudiera ser la misma que procese a la amidofosforribosiltransferasa de B. subtilis. Se ha especulado sobre la posibilidad de que el procesamiento que expone la cisteína en la enzima de B. subtilis, sea un paso necesario para que tal residuo pueda participar en la unión de glutamina. Es probable que esta propuesta sea cierta, si se considera que la enzima de E. coli también posee un residuo de cisteína en su extremo amino terminal. La necesidad de procesamiento en la enzima de B. subtilis puede reflejar un mecanismo que permita la regulación de la actividad de la enzima. Por otro lado, el procesamiento en la subunidad grande de GOGAT es un evento que no se había sospechado. Hasta el momento se desconoce cual pudiera ser el papel de tal procesamiento, aunque una posible explicación es que tuviera una función similar a la mencionada arriba para la enzima de B. subtilis. Es importante señalar que puesto que la GOGAT de E. coli es la única GOGAT que se ha secuenciado, no es posible en este momento saber si este

tipo de procesamiento se lleva a cabo en la GOGAT de otros organismos.

PERSPECTIVAS

La información que hasta la fecha se ha logrado obtener sobre el operón de GOGAT que codifica para las dos subunidades de esta enzima y una posible proteína regulatoria (Castaño et al., 1988), abre la posibilidad de plantear una serie de experimentos que permitan conocer con más detalle a este operón y su papel dentro de la regulación de la asimilación de amonio.

Es poco lo que se conoce sobre los factores que participan en la regulación de la expresión de GOGAT. La determinación de la secuencia nucleotídica de la región flanqueadora 5' para gluB y la identificación del promotor funcional para el operón bajo condiciones de crecimiento con alto amonio, han permitido proponer secuencias con un posible papel regulatorio.

Son tres los principales elementos que se proponen como posibles moduladores de la expresión de GOGAT, y a continuación se describen estrategias de estudio para analizar cada una de ellas:

- 1.- El papel de la secuencia "Box A", posiblemente involucrada en antiterminación dentro de la expresión del operón, puede ser estudiado utilizando cepas deficientes en antiterminación. Este estudio se vería facilitado por la obtención de fusiones de los genes para cada una de las dos subunidades con el gene que codifica para la beta-galactosidasa. La comparación entre la

expresión de las proteínas de fusión en la cepa silvestre y la cepa deficiente en antiterminación, permitiría identificar el papel de ese mecanismo en la expresión de gltB y gltD.

2.- La técnica de transcripción in vitro puede ser útil para demostrar si el palíndromo localizado en la región flanqueadora 5' tiene o no un papel como terminador de la transcripción. Con esta técnica es posible analizar los tamaños de los transcritos generados a partir del promotor identificado y de esta manera localizar el sitio de término de la transcripción (Mahadevan y Wright, 1987).

3.- La posibilidad de que pueda existir un fenómeno similar al de atenuación relacionado con la regulación de GOGAT, dependería en primer lugar de que se traduzca la secuencia comprendida en la posición 198-228 (Fig. 2, Oliver et al., 1987). Una forma de establecer que tal región codifica para un péptido, puede ser fusionar esa región con un gene que carezca de regiones de inicio de la traducción. La producción de tal proteína dependerá de que las secuencias de inicio de la traducción sean funcionales en el gene que se está analizando. Tal fusión se puede hacer con el gene modificado de la beta-galactosidasa.

Si se demuestra la funcionalidad de una o varias de las regiones propuestas, entonces se deberá seguir con su caracterización. Esto puede realizarse a través de experimentos de mutagénesis dirigida, en los que se modifiquen las secuencias que se consideren importantes. También, dependiendo de que

regiones se demuestre son funcionales, se podrá proponer un mecanismo de acción en el que se tome en cuenta la acción conjunta de esos elementos.

Los sitios de unión a substratos y a cofactores que se han propuesto a partir del análisis de las secuencia de aminoácidos de las dos subunidades, deben de ser demostrados experimentalmente. Una estrategia sería aplicar la técnica de mutagénesis dirigida para cambiar aminoácidos específicos en la proteína. Las subunidades mutantes resultantes podrían ser analizadas comparando sus afinidades por la molécula cuyo probable sitio de unión fué cambiado, con la correspondiente de la subunidad silvestre. En el caso de sitios con un posible papel modulador de la actividad catalítica, las mutantes respectivos podrían ser evaluados in vitro, comparándolas con la enzima silvestre. Los genes para dichas mutantes, podrían ser introducidos a E. coli para de esta manera poder observar el fenotipo conferido por los cambios en la secuencia polipeptídica.

Se podría pensar también en la construcción de proteínas híbridas, substituyendo algunos de los posibles dominios funcionales de GOGAT, por dominios ya caracterizados en otras proteínas. Un ejemplo es substituir la región que incluye a los dos posibles sitios unión a dinucleótido en GOGAT, por la región equivalente de una de las enzimas con las que muestra mayor similitud. Esto permitiría evaluar si los sitios provenientes de GOGAT son funcionales en otra proteína. Aunque es importante mencionar que este tipo de experimentos requieren de un

conocimiento muy detallado de las estructuras de las proteínas en cuestión.