

03088

3
2ej

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Instituto de Biotecnología

**Caracterización, manipulación y posible papel en la Naturaleza
del gene *pac* y su producto, la enzima penicilino acilasa de
Escherichia coli.**

TESIS

Que para obtener el grado de
Doctor en Biotecnología

presenta

Enrique Merino Pérez

TESIS CON
FALLA DE ORIGEN

Cuernavaca, Morelos

1992

**TESIS CON
FALLA DE ORIGEN**



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.

Contenido

Capítulo I. Prólogo y Objetivos

Capítulo II. Antecedentes y Justificaciones

II.1. Antibióticos β -lactámicos, naturales y semisintéticos

II.2. Importancia de la enzima penicilino acilasa (PA)

II.2.1. Aspecto tecnológico.

II.2.2. Estado actual del conocimiento sobre el gene *pac* y de su producto, la enzima PA.

Capítulo III. Mecanismos y elementos involucrados en la regulación de la actividad de penicilino acilasa.

III.1. Introducción

III.1.1. Estructura del gene *pac* de *E. coli* y de sus regiones de regulación 5' y 3'.

III.1.2. Estructura de la enzima PA de *E. coli*.

III.2. Resultados y Discusión

III.2.1. Estudios sobre la regulación transcripcional del gene *pac*.

III.2.1.1. Estudios de fusión de los genes *pac-lacZ*.

III.2.1.2. Cuantificación del mRNA específico del gene *pac*.

III.2.1.4. Análisis de la región 5' de regulación del gene *pac*.

III.2.1.5. Presencia del gene *pac* en cepas de *E. coli*.

III.2.2. Estudios de la regulación post-transcripcional de la actividad de PA.

III.2.2.1. Determinación de la actividad específica de PA en extractos celulares.

III.2.2.2. Estudio de proteínas híbridas de PA.

III.3. Conclusiones particulares

Capítulo IV. Incremento de la actividad de la enzima penicilino acilasa mediante técnicas de Ingeniería Genética

IV.1. Introducción

IV.2. Resultados y Discusión

IV.2.1 Construcción de vehículos moleculares para lograr la sobre-expresión del gene *pac*

IV.2.2. Estudios de procesos fermentativos de cepas con actividad de PA.

IV.2.2.1. Efecto de la concentración y del tiempo de adición de inductor del sistema en la actividad de PA.

IV.2.2.2. Efecto del fondo genético sobre la actividad de PA.

IV.3. Conclusiones particulares

Capítulo V. Posible papel de la enzima penicilino acilasa en la Naturaleza

V.1. Introducción

V.1.1. Distribución, propiedades y mecanismo de acción de la enzima PA.

V.1.2. Vía de utilización del ácido fenilacético en *Pseudomonas* y *E. coli*.

V.2. Resultados y Discusión

V.2.1. Crecimiento de cepas de *E. coli* en presencia de compuestos fenilados.

V.2.2. Posible papel fisiológico de la enzima PA.

V.3. Conclusiones particulares

Capítulo VI. Conclusiones generales y perspectivas

Capítulo VII. Otros estudios realizados durante el Doctorado

- Origen y evolución de la transmisión de la información genética
- Estudio de genes sobrelapados y sus implicaciones evolutivas

Bibliografía

Índice de Tablas y Figuras

Nomenclatura

Apéndice I.- Artículos en los que se incluyen elementos de esta Tesis

- The role of penicillin amidases in Nature and in industry.
Trends in Biochemical Sciences.
- Carbon regulation and the role in Nature of the *E. coli* penicillin acylase (*pac*) gene.
Molecular Microbiology.
- A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC/M13 vectors.
BioTechnics.
- Recovery of DNA from agarose gels stained with Methylene Blue.
BioTechnics.

Apéndice II.- Otros artículos elaborados durante el período Doctoral

- New insights on the Comma-less theory.
Origins of Life.
- Are overlapping open reading frames remnants of a primaeval genetic code system?
Manuscrito en preparación.

Capítulo I. Prólogo y Objetivos

La enzima penicilino acilasa (PA) (E.C. 3.5.1.11) es la enzima industrialmente usada para hidrolizar a la penicilina G (PenG) en ácido fenilacético (AFA) y ácido 6-amino penicilánico (6-APA). Este último es un intermediario en la producción de penicilinas semisintéticas, que son en nuestros días, el tipo de antibiótico más importante para el tratamiento de enfermedades infecciosas. La enzima PA presenta características únicas dentro de las enzimas de origen procarionte; la forma activa de la enzima corresponde a la de un heterodímero, producto del procesamiento de un precursor común.

Paralelo a la importancia de esta enzima, el estudio del gene que la codifica, *pac*, representa un excelente modelo para conocer, a nivel molecular, algunos de los mecanismos y elementos involucrados en la expresión genética, ya que se encuentra modulado a través de diferentes sistemas de regulación. Por otro lado, el comprender la organización y la regulación de este gene, es un aspecto importante para lograr la obtención de cepas microbianas, modificadas genéticamente, que permitan sobreproducir la actividad de PA. Por estas razones, el gene *pac* y su enzima PA, representan un modelo de estudio de gran importancia, tanto en el campo básico como en el tecnológico.

En base a lo mencionado, la presente Tesis de Doctorado ha tenido como Objetivo General el obtener información sobre los mecanismos y elementos involucrados en la regulación de la enzima PA, y de como este conocimiento a través de la manipulación de su gene estructural *pac*, permita obtener cepas bacterianas sobreproductoras de la actividad de PA. Así mismo es también parte del Objetivo General de esta Tesis, el elucidar el posible papel fisiológico de la enzima PA en la Naturaleza.

Así mismo, los objetivos particulares de esta Tesis son:

- Integrar la información publicada más relevante en torno de las enzimas PA y de los genes estructurales que las codifican en diferentes organismos y especialmente en *E. coli*.
- Estudiar algunos de los mecanismos y elementos más importantes involucrados en la regulación de la actividad de PA: a) Organización y regulación del gene *pac*. b) Estructura de la PA y de su regulación post-transcripcional.
- Obtener cepas de *E. coli*, sobreproductoras de la actividad de PA.
- Proponer el posible papel de la enzima PA en la Naturaleza.

El contenido de esta Tesis esta organizado en capítulos, y en cada uno de ellos se trata en forma particular, los objetivos planteados.

Finalmente, se incluyen en el Apéndice I, los artículos elaborados durante el período Doctoral: Los dos primeros artículos están directamente relacionados con la Tesis Doctoral. Los dos segundos artículos son referentes a innovaciones metodologías, desarrolladas para facilitar parte del estudio experimental de la Tesis. Los dos últimos artículos fueron realizados paralelamente a esta Tesis, y están integrados en otra línea de investigación de nuestro laboratorio que tiene como objetivo estudiar el origen y la evolución de los primeros mecanismos de transmisión de la información genética.

Capítulo II. Antecedentes y Justificación

II.1. Antibióticos β -lactámicos, naturales y semisintéticos.

El descubrimiento de los agentes antimicrobianos marcó el advenimiento de una nueva era en la práctica médica. Desde entonces, el control de enfermedades infecciosas se ha basado en la elección de un gran número de compuestos inhibitorios del crecimiento bacteriano a los que se les ha asignado el nombre genérico de antibióticos. Desde el descubrimiento de la penicilina por Fleming en 1929, cientos de antibióticos han sido descubiertos o generados por vías semisintéticas. Así, para 1981, se habían desarrollado cerca de 20,000 penicilinas, 4,000 cefalosporinas, 1,000 rifampicinas, 500 cloranfenicoles, 500 kanamicinas y 250 tetraciclinas. En ese mismo año, la producción de antibióticos alcanzó aproximadamente las 25,000 toneladas, de las cuales cerca de 12,000 correspondieron a penicilinas, 5,000 a tetraciclinas, 1,200 de cefalosporinas y 800 de eritromicinas.^{25,73} Del anterior grupo de antibióticos, los β -lactámicos (penicilinas y cefalosporinas), se caracterizan por poseer un anillo β -lactámico en su molécula (Fig. 1). La diversidad de este tipo de compuestos se da a partir de la modificación de su núcleo básico (Fig. 2). Este grupo confiere al antibiótico la capacidad para interactuar covalentemente con proteínas periplásmicas (PBPs) involucradas en la síntesis de pared celular bacteriana en sus etapas terminales.⁹⁴

Las cepas usadas industrialmente para incrementar los niveles de producción de estos antibióticos han sido obtenidas a partir de programas de desarrollo a gran escala. Por ejemplo, durante 40 años, la compañía Gist-Brocades (líder actual en la producción de penicilinas), ha mejorado la productividad de su proceso en 1,000 veces.⁶⁸ Sin embargo, después de la década de los 60s, el uso desmedido de los antibióticos β -lactámicos en la terapia clínica así como su aplicación en otras áreas tales como la preservación de alimentos y la nutrición animal, propiciaron la selección de microorganismos resistentes a los mismos.

El mecanismo más común de resistencia bacteriana a estos antibióticos, es su inactivación mediante la ruptura del enlace amídico del anillo β -lactámico por las enzimas β -lactamasas.⁷² Actualmente, estas enzimas constituyen uno de los más serios problemas de la terapia clínica, ya que se presentan en una gran variedad de bacterias patógenas.

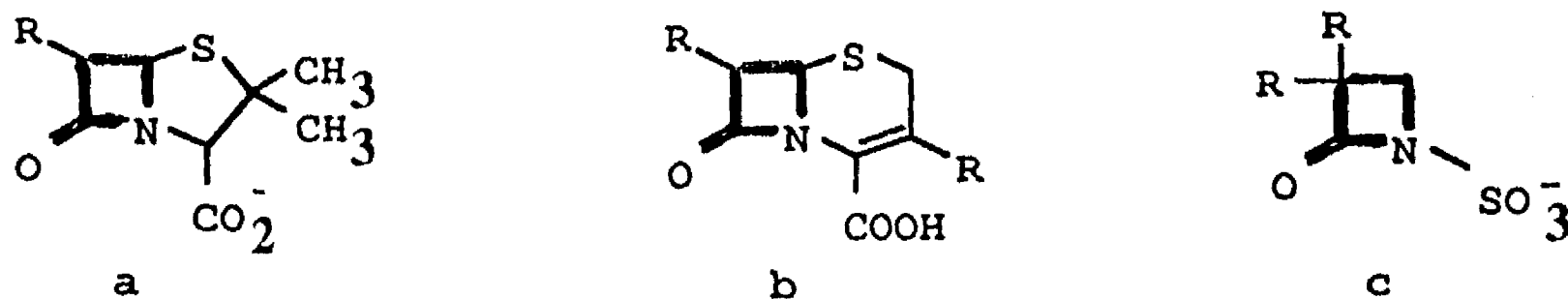
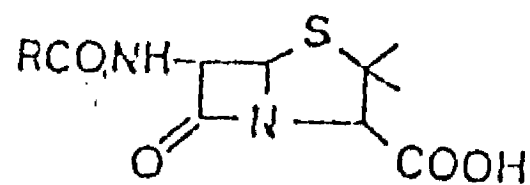


Figura 1. Estructura general de un antibiótico β -lactámico.

a) penicilinas, b) cefalosporinas, c) monobactamas. El grupo β -lactámico se indica con las líneas gruesas.

El problema ha sido parcialmente resuelto con el uso de nuevos antibióticos β -lactámicos semisintéticos con estructuras modificadas, que son substratos pobres para estas enzimas, a comparación de los antibióticos naturales. Sin embargo, se ha encontrado que cambios puntuales en los genes de las β -lactamasas, pueden modificar el espectro de su especificidad, por lo que es claro que ningún antibiótico β -lactámico podrá permanecer en el mercado por tiempo indefinido.^{88,47,53,89,74}

Las penicilinas producidas por vía semisintética, son en la actualidad los antibióticos que dominan el mercado por ser, no tan solo los más baratos, sino también los menos tóxicos. Paralelamente a lo anterior, es importante mencionar que el empleo en la clínica de un nuevo y potente inhibidor de β -lactamasas, el ácido clavulánico, combinado con las penicilinas semisintéticas existentes, permite rescatar en muchos casos, el uso de varias de ellas.¹⁴



R Nombre

R Nombre

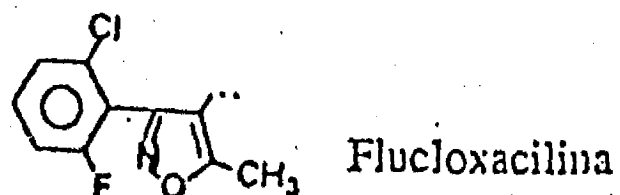
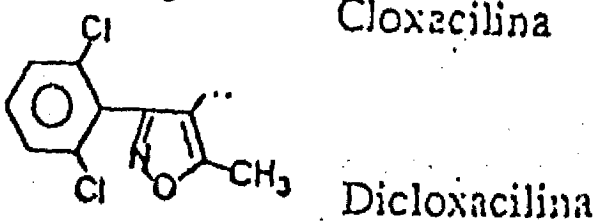
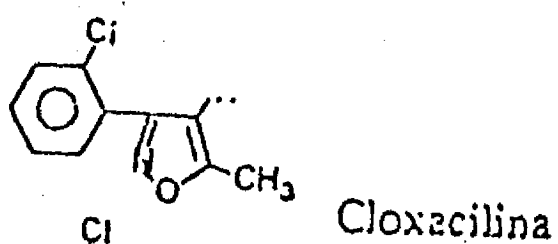
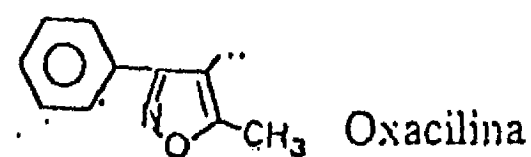
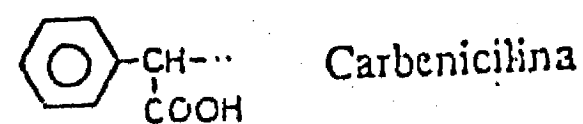
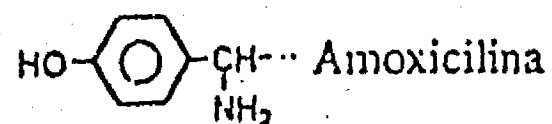
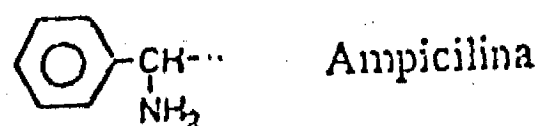
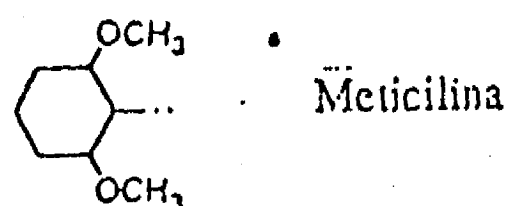
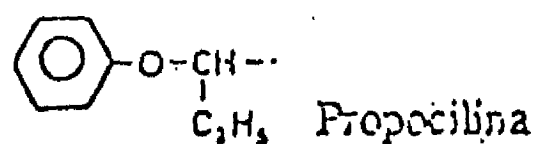


Figura 2. Estructura química de algunas penicilinas semisintéticas. En la parte superior de la figura se muestra la estructura general de las penicilinas semisintéticas. En la parte inferior se muestra los diferentes tipos de radicales, que dan origen a las correspondientes penicilinas semisintéticas.

II.2. Importancia de la enzima penicilino acilasa.

II.2.1. Aspecto tecnológico.

El primer reporte de la obtención del 6-APA de un proceso fermentativo data del año de 1959, cuando Batchelor y colaboradores lograron obtenerlo utilizando una cepa de *Penicillium chrysogenum*.⁶ A partir de entonces, fue posible la elaboración de penicilinas semisintéticas, y con ello se logró una gran revolución en la práctica terapéutica. Dificilmente se puede pensar en ejemplos similares, en donde la utilización de una enzima sea responsable de una contribución de tal importancia en el campo de la medicina.

Pese al su gran éxito, la baja producción y los complejos mecanismos involucrados en el aislamiento del 6-APA, hizo del proceso desarrollado por Batchelor y colaboradores, un proceso poco rentable en aquel momento. Hoy en día, el 6-APA es producido por hidrólisis química o enzimática de las penicilinas G y V. Actualmente, la producción vía enzimática del 6-APA es preferida sobre la realizada mediante procedimientos químicos, ya que además de resultar más barata, las sustancias utilizadas en el proceso químico son altamente contaminantes, y por lo tanto de uso cada vez más restringido.⁶⁹

El proceso enzimático para la obtención de 6-APA se basa, como ya se ha mencionado, en la hidrólisis de la PenG utilizando a la enzima PA. Los productos de esta reacción son el AFA y 6-APA (Fig. 3). Este último al ser reacilado en el α -amino de la posición 6 con una cadena lateral particular, permite producir una gran variedad de compuestos β -lactámicos^{58,64} (Fig. 2 y 3).

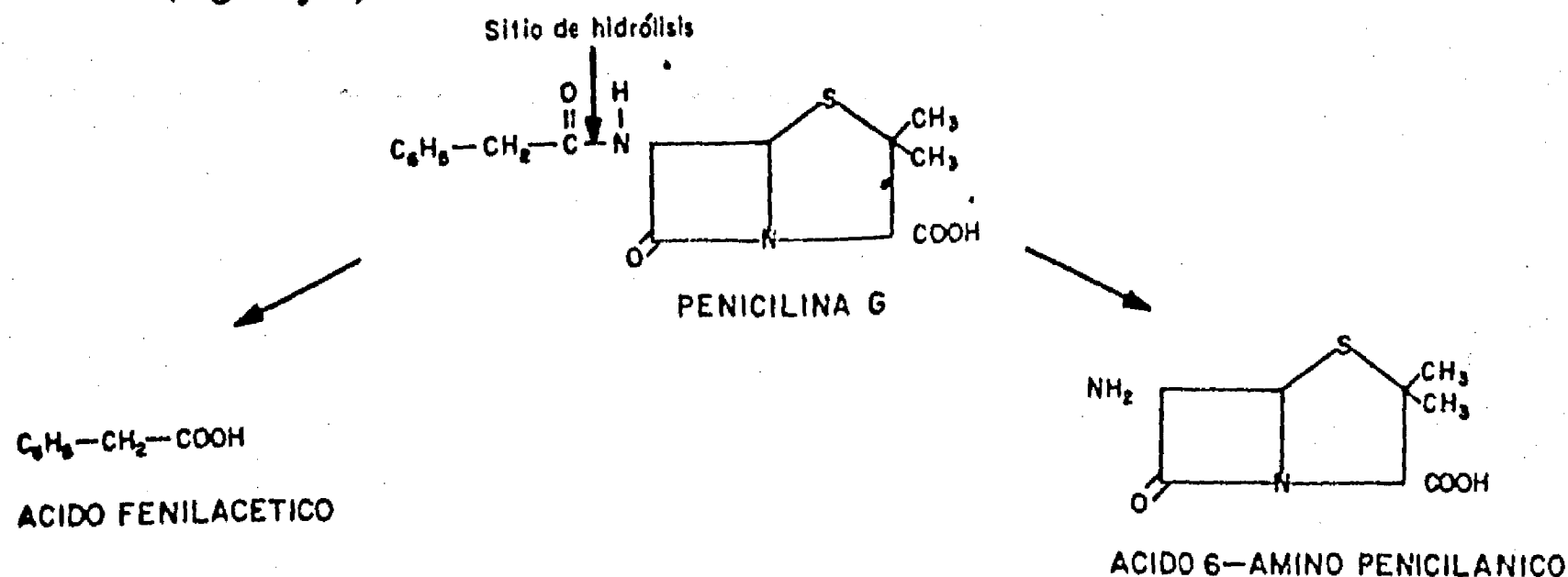


Figura 3. Esquema de la acción de la enzima penicilino acilasa sobre la molécula de penicilina G. En la figura se muestra el sitio de hidrólisis de la PenG por la enzima PA, así como los productos de esta hidrólisis, AFA y 6-APA.

Aunado a lo anterior, el desarrollo de la tecnología de inmovilización de enzimas ha permitido la elaboración de biocatalizadores en los que la enzima en cuestión puede ser reusada en varias ocasiones. La producción industrial de 6-APA a través de la inmovilización de la enzima PA, ha sido uno de los primeros éxitos en la aplicación comercial de esta tecnología.⁵⁸

II.2.2 Estado actual del conocimiento sobre el gene *pac* y de su producto, la enzima PA

En *E. coli*, la forma activa de esta enzima consiste de dos subunidades heterólogas producto del procesamiento proteolítico de un precursor común^{8,9,71} (Fig. 4). Esta forma de activación es similar a la que se realiza en algunas hormonas de tipo eucariote, como la insulina, o en algunos zimógenos como el fibrinógeno o el tripsinógeno; sin embargo, se puede considerar que entre los organismos procariotes, este tipo de procesamiento de las enzimas capaces de deacilar penicilinas o cefalosporinas, como son los casos de la PenG acilasa de *E. coli*,^{8,9,71}, *Arthrobacter viscosus*,⁶¹, *Bacillus megaterium*,⁷⁹, *Proteus rettgeri*,^{11,12,21,24} *Kluyvera citrophila*,^{5,29,76} y la enzima GK16 glutaril 7-ACA acilasa,⁴⁴ y la enzima SE83 cefalosporina C acilasa⁴⁵ de *Pseudomonas melanogenum*, es poco frecuente.

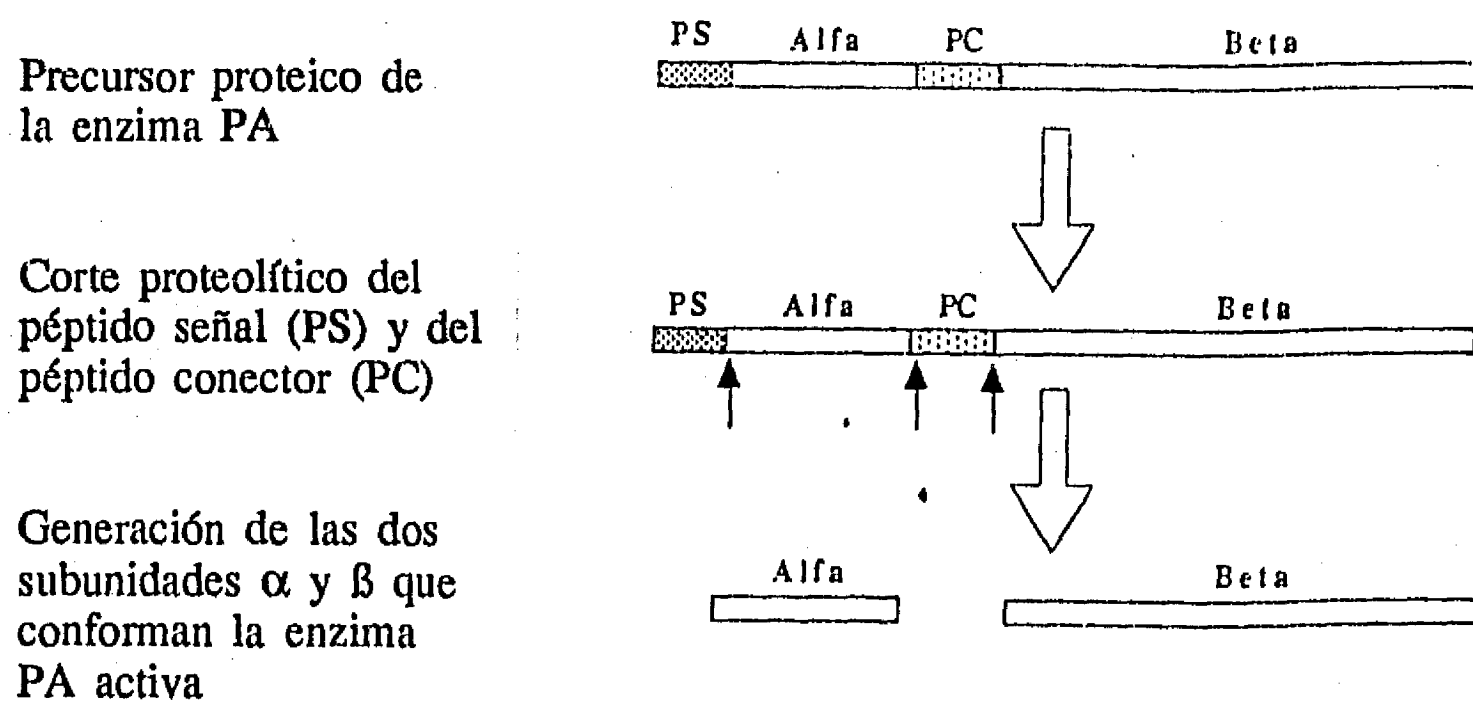


Figura 4. Ruta del procesamiento del precursor proteico de PA. Alfa y Beta representan a las subunidades proteicas de 24 y 62 kDa respectivamente. PC representa al péptido conector y PS al péptido señal. La forma activa de la enzima PA corresponde a las subunidades alfa y beta correctamente asociadas. Ninguno de los intermediarios del procesamiento presenta actividad de PA.⁷¹

Ligado a lo anterior, es importante señalar que la expresión del gene *pac*, es un excelente modelo para conocer, a nivel molecular, algunos de los elementos involucrados en la expresión genética de este gene y finalmente de su producto proteico. La actividad de PA se encuentra modulada por diversos elementos, siendo los más importantes la inducción por AFA, regulación por fuente de carbono y temperatura de crecimiento.

Inducción por ácido fenilacético: El AFA es uno de los dos productos de la hidrólisis de la PenG por la enzima PA (Fig. 3). Se ha reportado que la actividad de PA en células crecidas en AFA (0.1 a 0.3% p/v), es aproximadamente 8 a 10 veces mayor que la que se obtiene al crecer en ausencia de este compuesto.⁸³ Esta inducción es altamente específica del AFA, ya que muchos compuestos estructural y químicamente relacionados al mismo, no producen ningún efecto.⁹² Se ha reportado que la actividad de PA, tanto en *Proteus rettgeri*,^{21,24} como en *Kluyvera citrophila*,²⁹ no es inducida por AFA. No obstante, cuando el gene *pac* de *K. citrophila* es clonado en *E. coli*, el efecto de inducción por este compuesto puede ser observado.²⁹ Es importante mencionar también que el AFA es un inhibidor del crecimiento celular y que concentraciones superiores a 0.3%, son tóxicas para las bacterias.¹⁸

Un estudio más detallado de esta regulación, así como las contribuciones en este respecto, realizadas durante la presente Tesis, se incluyen en el artículo "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", que se anexa en Apéndice I.

Regulación por fuente de carbono: Se ha reportado que la expresión del gene *pac* de *E. coli* se encuentra bajo el control de represión catabólica por diferentes fuentes de carbono, tales como glucosa, lactosa, fructuosa y glicerol.^{28,93,94,83} Esta represión puede ser totalmente liberada mediante la adición exógena de AMPc.²⁸ Hasta el momento de iniciar la presente Tesis, no había sido demostrado si el efecto de represión catabólica se ejercía directamente sobre el inicio de transcripción del gene *pac*, o bien de una manera indirecta, mediante la regulación de otros genes cuyos productos estuvieran relacionados con el transporte del inductor AFA, o el transporte y/o procesamiento del precursor de PA.⁹³ El efecto, directo o indirecto, de la represión catabólica por glucosa sobre la actividad de PA en células crecidas en un medio que contiene glucosa y AFA es de aproximadamente 30% con respecto a aquellas células crecidas en ausencia de glucosa.⁹³

Es interesante mencionar que la actividad de PA en *P. rettgeri* es activada por glucosa y reprimida por succinato, fumarato y malato,²¹ y que cuando el gene *pac* de este organismo

es transformado en cepas de *E. coli*, no se observa ningún efecto de inhibición por succinato, fumarato o malato pero si se presenta una aparente inhibición catabólica por glucosa.²⁴

En el Apéndice I, se anexa el artículo "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", en donde se analiza en detalle este tipo de regulación, incluyendo las contribuciones realizadas durante la presente Tesis.

Termoregulación: La temperatura de crecimiento es un factor determinante en la actividad final de PA. Se ha reportado que la temperatura óptima de crecimiento de *E. coli* para obtener la actividad máxima de PA es 28 °C. A temperaturas mayores de crecimiento, la actividad de PA decrece, de tal manera que, en células crecidas a 37 °C, la actividad de PA es casi nula.⁹⁰ Esta pérdida de actividad de PA sin embargo, no es debida a la inactivación de la enzima por calor, ya que la temperatura óptima de reacción de esta enzima es de 42 °C.⁹⁰ Oh *et al.*, han demostrado que cepas crecidas a 42 °C transformadas con plásmidos multicopia que contienen al gene *pac* bajo el control del promotor PL del bacteriofago lambda, acumulan al precursor de PA no procesado.⁶⁰ A partir de este resultado, estos autores concluyen que la termoregulación de la actividad de PA esta mediada a nivel del transporte, y/o procesamiento del precursor de PA. Sin embargo, es necesario tener en cuenta que estas determinaciones fueron realizadas con un promotor diferente del nativo y de que de los experimentos realizados, no es posible inferir si tal acumulación es debida a un efecto de saturación del mecanismo de transporte y/o procesamiento originada por la sobre-expresión del precursor de PA, o bien por otro tipo de regulación.

Valle F. ha demostrado que la termoregulación de la actividad de PA es independiente del locus *envY*⁸⁶, cuyo producto esta involucrado en la termoregulación de genes de porinas *ompF*, *ompC* y *lamB*.⁴⁰

García y Buesa,²⁹ han reportado que el gene *pac* de *K. citrophila*, se termoregula de la misma manera que lo hace el gene *pac* de *E. coli*, no obstante el decremento de la actividad de la primera es tan solo de aproximadamente un 50% cuando es crecida a 37°C. La magnitud de esta regulación es independiente de que el gene *pac* de *K. citrophila* se encuentre en *K. citrophila*, o haya sido clonado en *E. coli*. Es interesante mencionar que Daumy *et al.* han reportado que la actividad de PA de *P. rettgeri* solo existe a 28 °C, ya que este organismo no es capaz de crecer a 37 °C.²⁴

Capítulo III. Mecanismos y elementos involucrados en la regulación de la actividad de penicilino acilasa.

III.1. Introducción

III.1.1. Estructura del gene *pac* de *E. coli* y de sus regiones de regulación 5' y 3'.

El gene estructural *pac* de *E. coli* ATCC1105 consta de 2,532 pb. En nuestro laboratorio hemos identificado, mediante experimentos de extensión de primer, que el inicio la transcripción del RNA mensajero se encuentra a 30 pb hacia arriba del primer codón ATG.⁸⁵ El análisis de la secuencia nucleotídica de la región 5' de regulación, ha permitido determinar mediante criterios de secuencia consenso, dos posibles sitios de unión para el complejo CRP-AMPc, tres secuencias palindrómicas que pudieran estar involucradas en la regulación transcripcional del gene, así como un promotor funcional 34 pb hacia arriba del gene estructural⁸⁵ (Fig. 5).

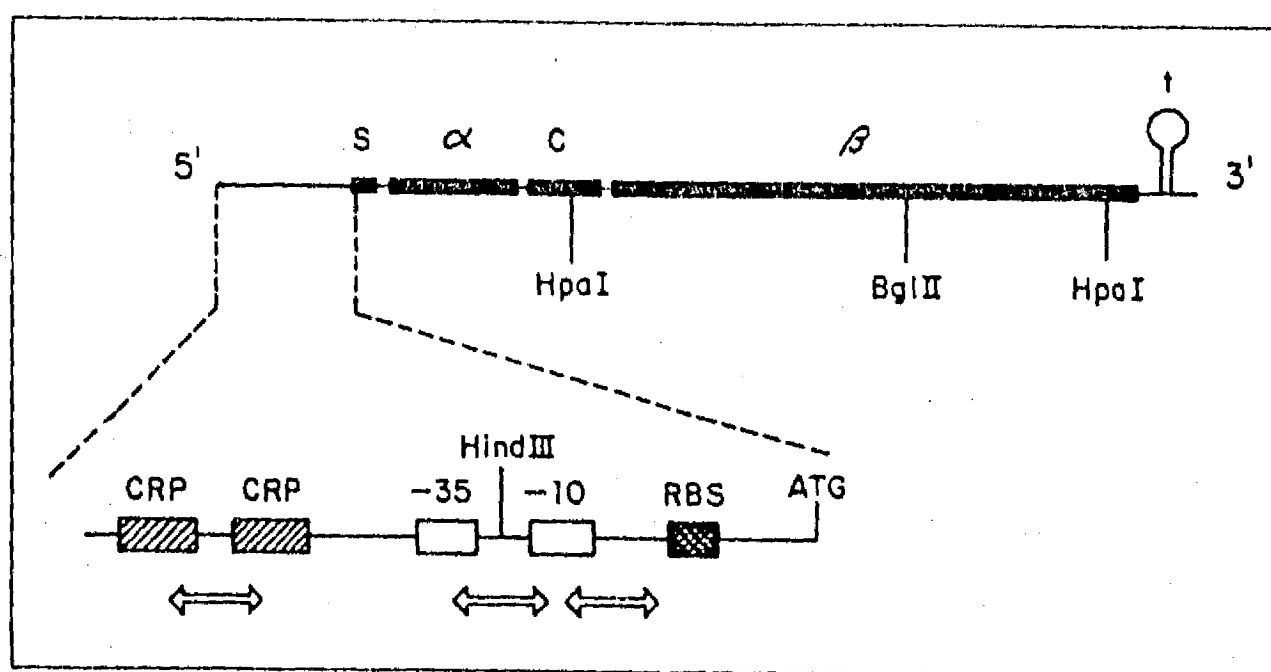


Figura 5. Organización física del gene *pac* de *E. coli*. En la figura se esquematiza al gene estructural *pac*, así como los elementos de regulación localizados en los extremos 5' y 3' terminales, estos son: dos sitios propuestos para el reconocimiento por el complejo CRP-AMPc (*CRP*), un promotor funcional (cajas -35 y -10), regiones palindrómicas imperfectas, posiblemente involucradas en la regulación de la expresión (flecha de dos cabezas), un sitio de unión a ribosoma (*RBS*) y el codón de inicio de la traducción (*ATG*). La estructura de tallo y asa (*t*) ubicada en el extremo 3', representa un posible terminador de la transcripción rho-dependiente⁸⁵.

La determinación del inicio de la transcripción del gene *pac* (Valle *et al*⁶⁵), fue realizada en condiciones de máxima inducción (AFA 0.1% como única fuente de carbono). Es importante mencionar que se han reportado evidencias en favor de la existencia de un segundo promotor ubicado en la región de DNA delimitada por el sitio de restricción *HindIII* inmediato anterior al gene estructural.⁶⁰ En la sección de resultados del presente capítulo, así como en el artículo anexo: "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", se realiza un análisis de la región de control 5' y se evalúa la existencia de este segundo promotor.

III.1.2. Estructura de la enzima PA

Como se ha mencionado en el capítulo II, la forma activa de la enzima PA se localiza en el espacio periplásmico y está constituida por las subunidades α (24 kDa) y β (62 kDa), en proporciones equimolares.^{8,9} Estas subunidades son el producto del procesamiento del precursor PA, cuyo peso original es de 94 kDa. Durante el proceso de translocación a través de la membrana interna, es removido el péptido líder de la región amino terminal del precursor. Una vez en el periplasma, se efectúa un segundo y tercer corte proteolítico con lo que se extingue el péptido conector ubicado entre las subunidades⁷¹ (Fig. 4). Ninguno de los intermediarios del procesamiento presentan actividad de amidohidrolasa.⁹

En base a resultados de experimentos de reasociación *in vitro*, Lindsay y Pain³⁹ han propuesto que la forma activa de la enzima PA, es adquirida a través del plegamiento secuencial de las dos subunidades. La subunidad α puede plegarse independiente en una estructura globular compacta. Por el contrario, el plegamiento de la subunidad β es dependiente de la conformación previamente adquirida por la subunidad α .⁷⁸ La función del péptido conector es la de favorecer la adecuada interacción entre las dos subunidades, que posteriormente permanecerán unidas por interacciones que son principalmente de tipo hidrofóbicas.³⁹

Aún no han sido identificadas las enzimas involucradas en el procesamiento de la enzima PA. Se ha propuesto que la propia enzima PA pudiera tener propiedades proteolíticas que le permitieran autoprocesarse.^{12,11} Sin embargo, se ha reportado que son necesarias ciertas proteasas del periplasma para obtener la forma madura de PA.⁷¹ Muy probablemente, estas proteasas están involucradas en el procesamiento de otras proteínas distintas a PA, ya que cepas de *E. coli* carentes de la actividad nativa de PA, al ser transformadas con el gene

pac de otros organismos (*Arthobacter viscosus*,⁶² *Bacillus megaterium*,⁷⁹ *Kluyvera citrophila*²⁹ y *Proteus rettgeri*²⁴), fueron capaces de sintetizar, y procesar correctamente a la enzima PA.

Mediante mutagénesis generalizada sobre el gene *pac*, se ha identificado que la especificidad de la enzima hacia grupos fenilo radica en la subunidad α ,²² mientras que la subunidad β contiene un residuo de serina reactivo que puede ser inactivado por PMFS.¹²

Aunado a lo anterior, se ha pensado que un estudio comparativo entre las enzimas PA de diversos organismos, constituye una fuente importante de información para entender algunas de sus propiedades estructurales, cinéticas y regulatorias. En la Tabla I se compilan algunas de ellas.

Microorganismo	<i>E. coli</i>	<i>K. citrophila</i>	<i>P. rettgeri</i>	<i>A. viscosus</i>
Precursor (kDa)	95	93.5	80-90	
Precursor (a.a.)	846	844		
Subunidad α (kDa)	24	23	24.5	24
Subunidad β (kDa)	65	62	65	60
Enzima madura (kDa)	89	85	89.5	81
Péptido señal (a.a.)	26	26		
Sitio de corte PS	ALA EQ	ALA AS		
Péptido conector (a.a.)	54	54		
Sitio de corte PC	TA AL TT SN	TA AL TT SN	SN	
Substrato	PenG	PenG Cefalexina Cefaloridina	PenG	PenG Cefalexina Cefalotina
K_M	20 mM	1.4 mM		0.42mM
V_{MAX}				62 μ M
pH óptimo de reacción	8.1	7.3		6 - 7.5
Temp. óptima de reacción		55 °C		45 - 50 °C
Inhibición por	AFA 200 mM 6-APA 1 mM			AFA 4 mM 6-APA 1 mM
Localización	Periplasma	Periplasma	Intracelular	Extracelular
Represión por	Glucosa	Glucosa	Succinato malato fumarato	
Temp. crecimiento	28 °C	30 °C	30 °C	28 °C

Tabla I. Propiedades estructurales, cinéticas y regulatorias de la enzima PA de diferentes microorganismos.

III.2. Resultados y Discusión

III.2.1. Estudio sobre la regulación transcripcional del gene *pac*

Como se mencionó anteriormente, debido a que la enzima PA requiere ser transportada y procesada en el espacio periplásmico para obtener su forma activa, la biosíntesis del precursor de la enzima PA, no necesariamente es un reflejo de la actividad final de la enzima. Con el objeto de discernir en que medida esta actividad es modulada a nivel del inicio de la transcripción del gene *pac*, se realizaron los siguientes estudios:

a) **Regulación por fuente de carbono:** En el artículo "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", se describe en forma detallada los elementos relacionados a este tipo de regulación. En dicho estudio se analizaron condiciones de exceso (11mM glucosa) y limitación (22mM glicerol) de carbono.

b) **Inducción por AFA:** Los elementos relacionados a este tipo de regulación se describen en forma detallada en el artículo antes mencionado. Los experimentos realizados en el estudio fueron llevados a cabo, tanto en presencia (AFA 0.1%), como en ausencia del inductor.

c) **Termoregulación:** El estudio de la termoregulación se efectuó al crecer las cepas a 28 y 37 °C, en condiciones de máxima inducción (AFA 0.1% como única fuente de carbono).

		Temperatura		Inductor
		28 °C	37 °C	
Fuente de carbono y energía	Glucosa 11 mM	A	Sin AFA	
		B	AFA 0.1%	
	Glicerol 22mM 22mM	C	Sin AFA	
		D	AFA 0.1%	
	-----	E	F AFA 0.1%	

Tabla II. Condiciones analizadas en el estudio de la regulación transcripcional del gene *pac*.

En la Tabla II se muestran las distintas condiciones de crecimiento utilizadas en los estudios anteriores. Para cada uno de las condiciones anteriores de estudio se realizaron las siguientes determinaciones:

III.2.1.1. Estudios de fusión de los genes *pac-lacZ*.

En este estudio se realizó la fusión de la región de regulación del extremo 5' terminal del gene *pac* con el gene reportero *lacZ*, que codifica para la enzima β -galactosidasa. Del plásmido pPA25⁶⁵ se aisló un fragmento de DNA que contiene la región de regulación del extremo 5' terminal del gene *pac* de *E. coli*, así como la parte del gene estructural correspondiente al péptido señal y a los primeros 75 aminoácidos de la subunidad α de la enzima. Este fragmento fue clonado en fase, al gene que codifica para la enzima β -galactosidasa en el plásmido pMC1403.¹⁵ El plásmido producto de esta clonación fue designado como pPA50. Las cepas de *E. coli* ATCC11101, así como la cepa HB101 transformada con los plásmidos pPA4,⁶⁵ pPA25⁶⁵ y pPA50 (Fig. 6), fueron crecidas hasta fase exponencial tardía en medio mínimo NN suplementado con 15mM NH₄Cl como fuente de nitrógeno y glucosa 11mM, glicerol 22mM o AFA 0.1%, como fuente de carbono y energía, de acuerdo a las condiciones descritas en la Tabla II. La actividad de PA se determinó mediante el método reportado por Balasingham *et al.*, en donde la formación de la base de Schiff colorida formada entre el p-dimetil aminobenzaldehído y el grupo amida del 6-APA, es una medida indirecta de la hidrólisis de la PenG.³ La actividad de β -galactosidasa fue determinada de acuerdo al método descrito por Miller, en base a la tasa de hidrólisis del ONPG⁵⁴.

III.2.1.2. Cuantificación del mRNA específico del gene *pac*.

El RNA total de la cepa ATCC11105 crecida en las diferentes condiciones anteriormente descritas, fue extraído de acuerdo al protocolo descrito por Young y Furano.⁹⁵ La integridad del RNA y el tamaño del mRNA de *pac*, fueron verificados en geles desnaturizantes de acrilamida 5%, urea 8M, y la pureza del mismo mediante estudios de tipo Northern⁸⁰ en geles de agarosa 1%, formaldehído 2.2M. La cuantificación del mRNA específico del gene *pac* fue realizada por hibridización en punto. La radioactividad de cada punto fue determinada en un contador de centelleo.

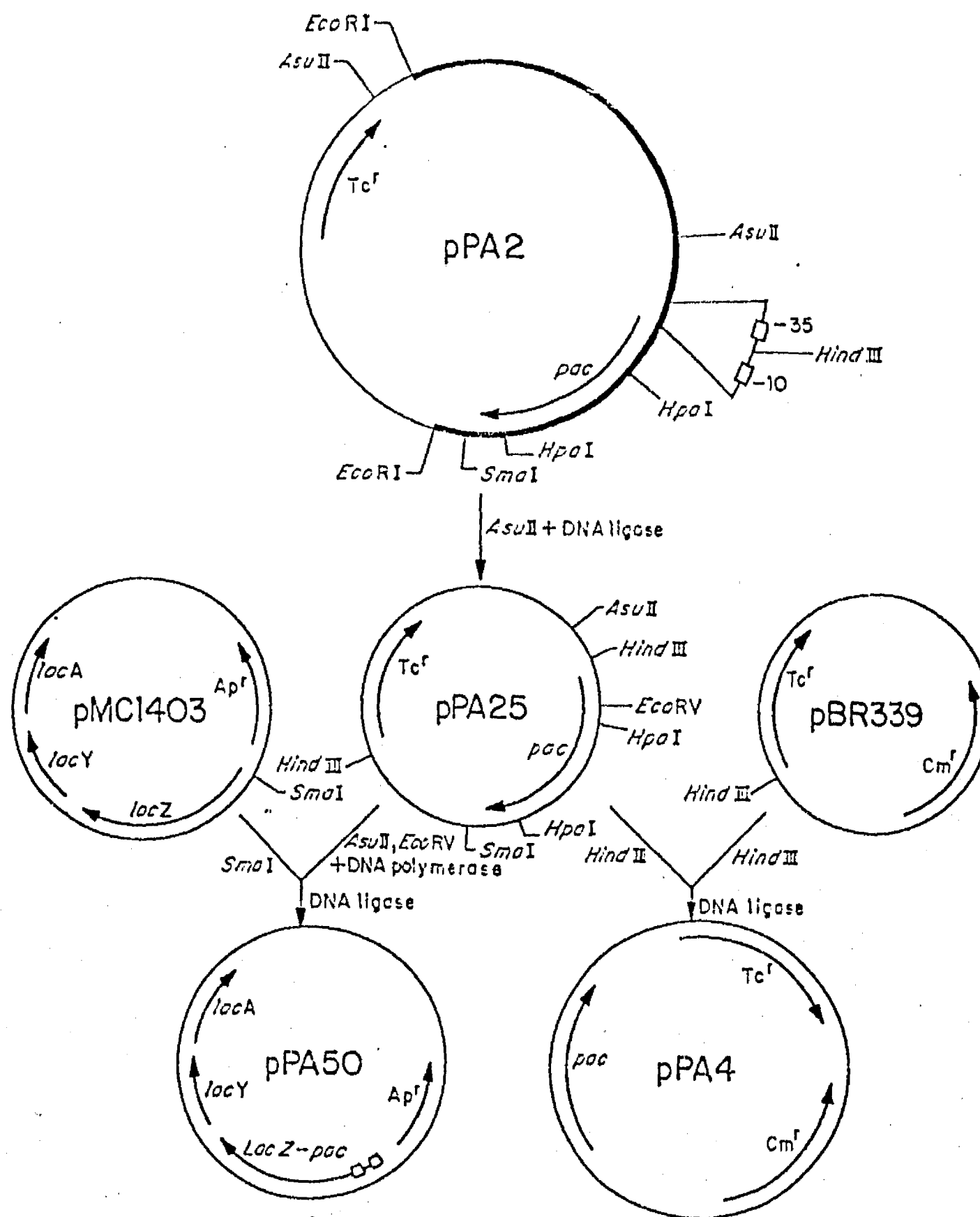


Figura 6. Construcción de vehículos moleculares para el estudio de la transcripción del gene *pac*. Los plásmidos pPA2, pPA4, pPA25, pPA25 y pBR339 han sido previamente descritos por Oliver *et al.*⁶⁵ El plásmido pMC1403 fue construido por Casadaban *et al.*,¹⁵ y contiene al gene reportero de *lacZ*. En plásmido pPA50, el gene *lacZ* ha sido fusionado a la región de regulación 5' del gene *pac*. La línea gruesa del plásmido pPA2 representa el fragmento cromosomal que contiene al gene *pac* de la cepa de *E. coli* ATCC11105.

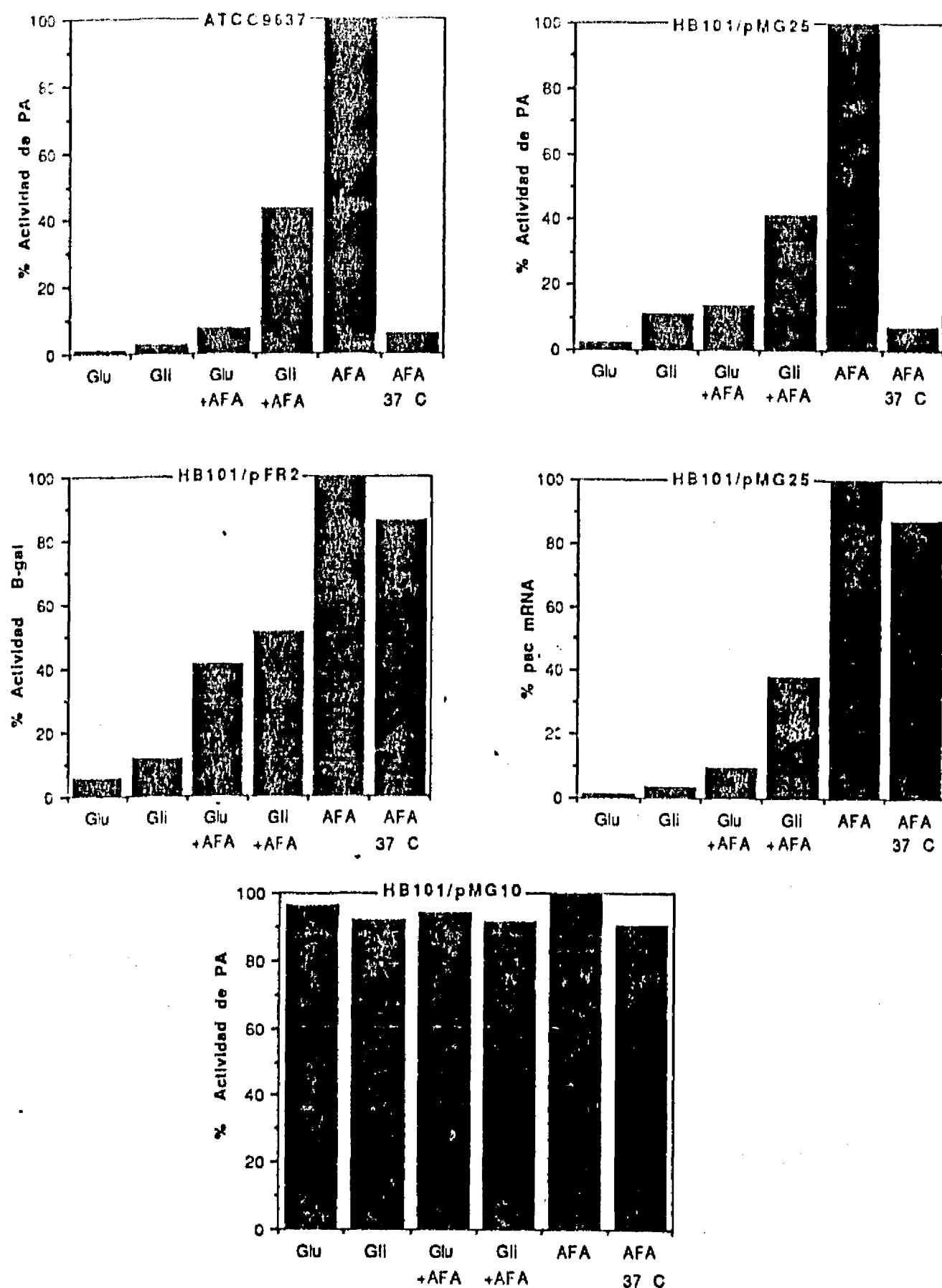


Figura 7. Determinación de las actividades de PA, β -galactosidasa y *pac* mRNA. Las cepas de *E. coli* ATCC11105 y HB101, transformadas con los plásmidos indicados, fueron crecidas en medio mínimo NN y cosechadas en la fase exponencial tardía. Seis condiciones fueron analizadas: **Glu:** glucosa 11 mM, 28 °C, **Gly:** glicerol 22 mM, 28 °C, **Glu+AFA:** glucosa 11 mM + ácido fenilacético 0.1%, 28 °C, **Gly+AFA:** glicerol 22 mM + ácido fenilacético 0.1%, 28 °C, **AFA:** ácido fenilacético 0.1%, 28 °C, **37 °C:** ácido fenilacético 0.1%, 37 °C. Con el objeto de poder realizar un análisis comparativo, los resultados son expresados en términos relativos al máximo valor obtenido (100%) de cada conjunto de determinaciones. El detalle y su discusión, de estos resultados se presentan en el artículo "Carbon regulation and the role in Nature of the *E. coli* penicillin acylase (*pac*) gene".

Los resultados correspondientes a la fuente de carbono e inducción por AFA, son descritos y analizados en forma detallada en el artículo: "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", y se resumen en la Fig. 7. Estos resultados muestran claramente que la regulación por represión catabólica, así como la inducción por AFA, es llevada a cabo a nivel del inicio de transcripción. Lo anterior está de acuerdo el reporte de Valle *et al.*,⁸⁵ en donde se propone la existencia de sitios para la unión al complejo CRP-AMPc en la región de control 5' de este gene.

Un elemento relevante de estos estudios, lo constituye los resultados de cuantificación de β -galactosidasa y del mRNA específico del gene *pac* en cultivos crecidos a diferentes temperaturas (Fig. 7). Estos resultados indican que la termoregulación de la actividad intracelular de PA, es esencialmente postranscripcional.

Es importante notar que el sistema de regulación por temperatura de la actividad de PA en *E. coli*, es diferente de los anteriormente descritos en la literatura: Regulación mediada por el producto de *envY*, que regula transcripcionalmente a los genes de ciertas porinas como OmpF, OmpB y LamB;^{86,40} al sistema de termoregulación por proteólisis mediado por el producto de *ompT*;^{86,57a} o al sistema global de regulación conocido con el nombre de "Heat shock".⁵⁹ Tomando en cuenta esta regulación postranscripcional de PA, y considerando que la enzima activa de PA no es labil a 37 °C, es posible pensar que entre los elementos que pudieran estar involucrados en esta termoregulación de PA se encuentran entre otros:

a) El transporte del precursor de PA al periplasma, b) El procesamiento del precursor de PA, c) La vida media de la enzima PA.

III.2.1.4. Análisis de la región 5' de regulación del gene *pac*.

Como se mencionó en la Introducción del presente capítulo, en nuestro grupo se ha identificado un promotor funcional del gene *pac*, 34 pb hacia arriba del primer ATG del gene estructural⁸⁵ (Fig. 5). Sin embargo, existe controversia en torno a la existencia de un segundo promotor funcional.⁶⁰ Con el objeto de evaluar la factibilidad de estos promotores, se desarrolló un programa para computadora basado en el algoritmo estadístico de Mulligan *et al.*,⁵⁷ con el que fueron analizados los 160 nucleótidos anteriores al gene estructural de *pac*. El número de posibles promotores identificados en esta región, depende del valor filtro usado en el análisis. Los resultados del estudio se muestran en el histograma de la Fig. 8,

en donde se puede observar que el promotor propuesto por nuestro grupo (B), posee el valor más alto de filtro, y por tanto es estadísticamente el más probable dentro de esta región, mientras que por otro lado, el segundo promotor, propuesto por Oh *et al.*⁶⁰ (A), tiene un valor de filtro tan bajo, que con él también se pueden encontrar adicionalmente otros 13 posibles promotores. En el artículo "Carbon regulation and the role in Nature of the *E. coli* penicillin acylase (*pac*) gene", se presenta una descripción detallada de estos resultados y del algoritmo matemático empleado.

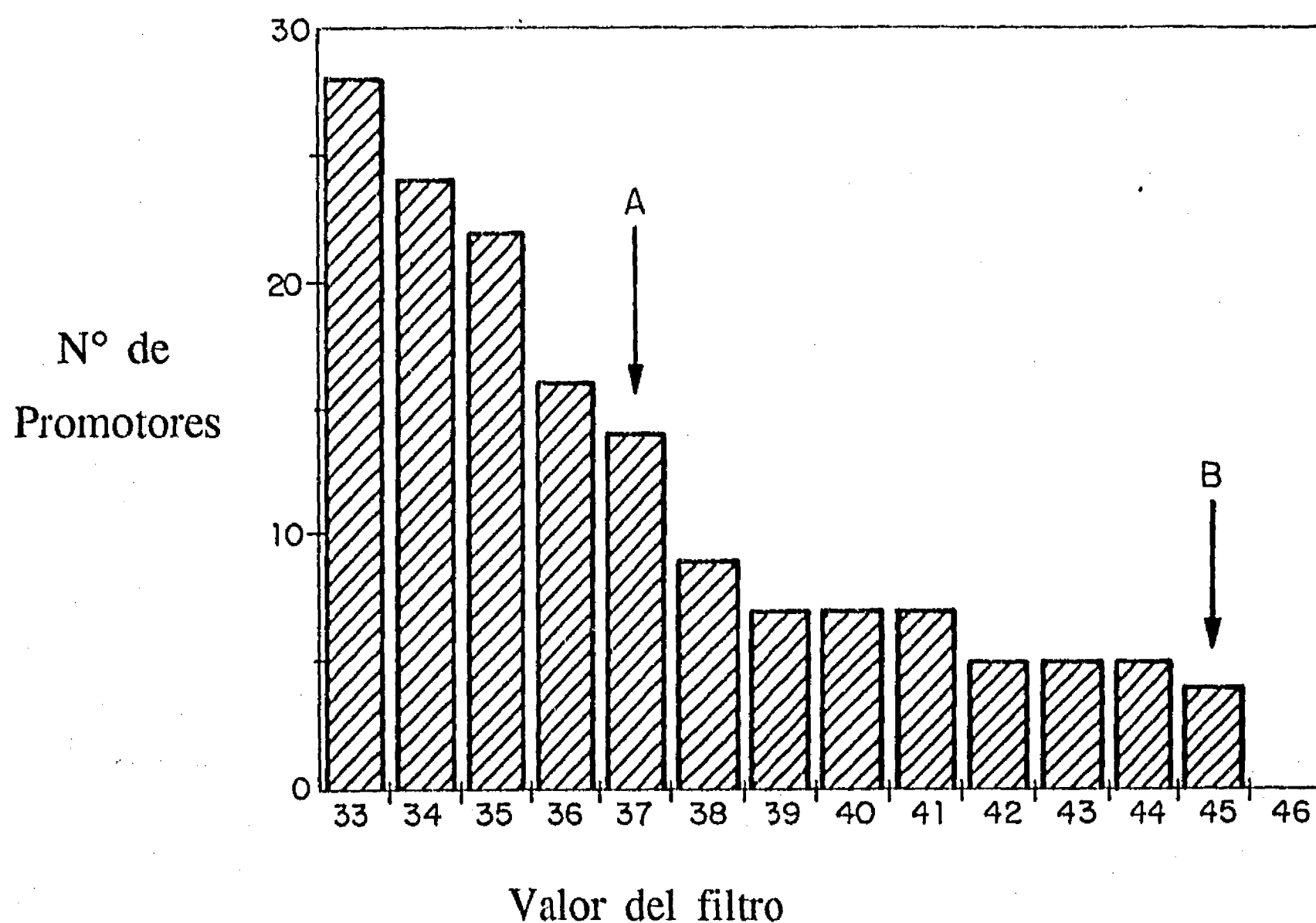


Figura 8. Análisis de los posibles promotores en la región 5' de regulación del gene *pac*. El histograma mostrado, representa el número de promotores σ^{70} identificados, dentro de los 160 nucleótidos anteriores al codón de inicio ATG, empleando el algoritmo estadístico propuesto por Mulligan *et al.*⁵⁷ Las flechas indican el valor del filtro correspondiente al promotor propuesto por Valle *et al.*⁸⁵ (B) y Oh *et al.*⁶⁰(A).

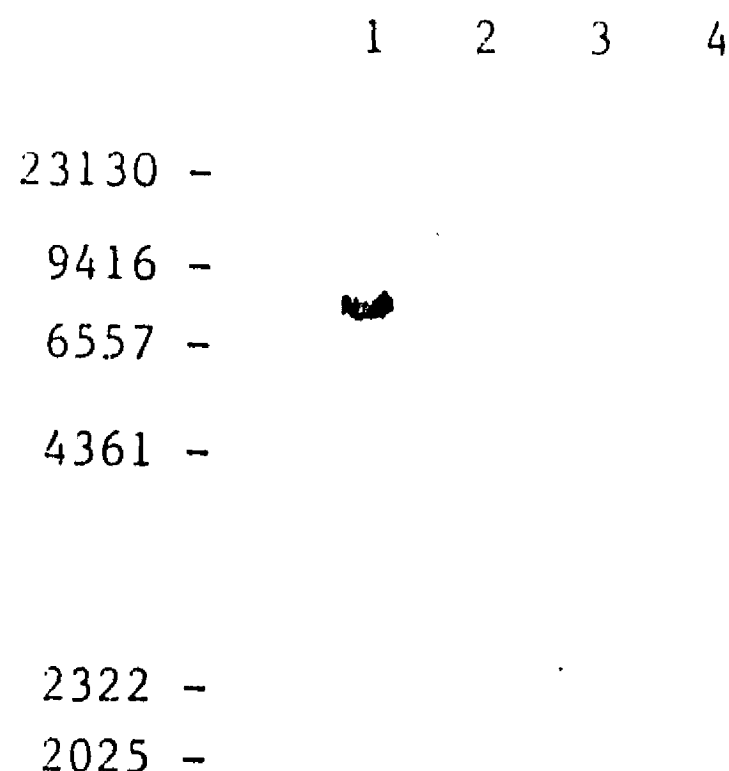


Figura 9. Detección del gene *pac* en cepas de *E. coli*. El DNA de las cepas de cepas de *E. coli* ATCC11105, HB101, JM101, W3110 (carriles 1, 2, 3 y 4, respectivamente), fueron digeridos con la enzima *EcoRI* y separados por electroforesis en gel de agarosa 1%, transferidos a una membrana de nitrocelulosa e hibridizados con el fragmento interno *HpaI*-*HpaI* del gene *pac*. El carril de la extrema izquierda corresponde a marcadores de peso molecular.

III.2.1.5. Presencia del gene *pac* en cepas de *E. coli*

Los estudios de inducción transcripcional del gene *pac* por AFA, revelaron que esta regulación puede estar presente en cepas de *E. coli* que originalmente carecían de actividad de PA. Este hecho puede implicar que: a) La regulación por AFA es común a otros genes presentes en *E. coli*, o b) El gene *pac* es un gene críptico en las cepas sin actividad de PA, pero la regulación por AFA en estas cepas es debida a que aún se encuentran presentes los elementos que anteriormente lo regulaban.

Con el objeto de discernir entre estas dos posibilidades, se realizaron experimentos tipo Southern⁸⁰ usando DNA cromosomal de cepas de *E. coli* que poseen (ATCC11105) o no (HB101, JM101, W3110) actividad de PA. La sonda usada en estos experimentos de hibridización, fue el fragmento interno del gene *pac* de 1.7 kb *HpaI*-*HpaI*. Como puede observarse en la Fig. 9, solo el DNA cromosomal de la cepa ATCC11105 presenta hibridización con el detector. Este resultado implica que el sistema que interviene en la regulación del gene *pac* por AFA en ciertas cepas de *E. coli*, es común a otros genes por lo que puede considerarse que el gene *pac* pudiera ser parte de un regulón modulado por AFA

III.2.2. Estudios de la regulación post-transcripcional de la actividad de PA.

III.2.2.1. Determinación de la actividad específica de PA en extractos celulares.

Como se demostró anteriormente, la actividad de PA se encuentra termoregulada esencialmente a un nivel post-transcripcional. Los mecanismos de regulación que se analizan en esta sección, se basan en la hipótesis de que la actividad de PA en células crecidas a 37 °C es nula debido a que a esa temperatura:

- a) No se produce la enzima responsable del procesamiento del precursor de PA.
- b) Se produce una proteasa(s) que degrada e inactiva a la enzima PA.

Con el objeto de analizar las anteriores alternativas, y en base a los resultados de Schumacher *et al.*,⁷² que demuestran que el precursor de PA acumulado en el citoplasma puede ser procesado después del rompimiento celular, se realizó el siguiente experimento: La cepa de *E. coli* HB101 (*pac*⁻) fue transformada con el plásmido pPA2. Posteriormente, ambas cepas fueron crecidas a 28 y a 37 °C, en condiciones de máxima inducción (AFA 1% como única fuente de carbono), cosechadas en fase logarítmica tardía, y resuspendidas en buffer fosfatos 0.1M, pH 7.5. Finalmente se determinó la actividad de PA a estos cultivos, así como al producto de la mezcla de ellos, tal y como se indica en la Tabla III.

Las condiciones 1 al 10, son controles del estudio en los que se observa que:

- a) La cepa de *E. coli* HB101 carece de actividad de PA, a menos que esta sea transformada con un plásmido que contenga al gene *pac*.
- b) La actividad de PA es casi nula en cepas de *E. coli* crecidas a 37 °C.
- c) El efecto de la ruptura celular mediante sonicado, no afecta la actividad de PA.
- d) La actividad específica de PA de la mezcla por partes iguales de un cultivo con actividad de PA con otro que no la tiene, es aproximadamente la mitad de la primera.

El resultado de la condición N°11 muestra que la disminución de la actividad de PA de las cepas crecidas a 37 °C, no es debida a que la endopetidasa(s) responsable del procesamiento del precursor de PA, sea sintetizada exclusivamente a 28 °C.

El resultado de la condición N°12 muestra que la termoregulación de la actividad de PA, no es debida a la inactivación de la enzima activa de PA, por proteasas que pudieran sintetizarse preferencialmente a 37 °C.

Cepa	Tratamiento del cultivo	Actividad relativa
1) HB101 37 °C	no sonicada	0%
2) HB101/pPA2 37 °C	no sonicada	4%
3) HB101 28 °C	no sonicada	0%
4) HB101/pPA2 28 °C	no sonicada	94%
5) HB101 37 °C	sonicado	0%
6) HB101/pPA2 37 °C	sonicado	6%
7) HB101 28 °C	sonicado	0%
8) HB101/pPA2 28 °C	sonicado	100%
9) HB101 37 °C + HB101/pPA2 37 °C	sonicadas e incubadas	3%
10) HB101 28 °C + HB101/pPA2 28 °C	sonicadas e incubadas	49%
11) HB101 28 °C + HB101/pPA2 37 °C	sonicadas e incubadas	2%
12) HB101 37 °C + HB101/pPA2 28 °C	sonicadas e incubadas	48%

Tabla III. Estudio del mecanismo de termoregulación de la actividad de PA. Con el objeto de poder realizar un análisis comparativo de los resultados, la actividad de PA se expresa en términos porcentuales respecto al máximo valor obtenido (100%).

III.2.2.2. Estudio de proteínas híbridas de PA.

Como se comentó anteriormente, el estudio comparativo del gene *pac* de diferentes organismos, puede brindar información relevante sobre la estructura y función de la enzima PA. En la presente Tesis se decidió realizar el estudio comparativo entre los genes *pac* de *E. coli* y *K. citrophila*. Pese a la gran homología a nivel de secuencia primaria entre estos dos genes (80%)⁵ (Fig. 10), su regulación, así como algunas propiedades enzimáticas de sus productos polipeptídicos presentan ciertas diferencias^{29,5,63,37,76}, siendo las más importantes las que presentan en su regulación por AFA y temperatura.

Con el propósito de correlacionar regiones específicas de la enzima PA, con algunas de las propiedades cinéticas o regulatorias de la misma, se decidió construir genes híbridos formados a partir de los genes *pac* nativos de *E. coli* y *K. citrophila*. El gene *pac* de *E. coli* utilizado en este estudio, había sido anteriormente aislado en nuestro laboratorio.⁶⁵ Con el objeto de aislar el gene *pac* de *K. citrophila*, el DNA cromosomal de este organismo fue

digerido con la endonucleasa de restricción *Hind*III y ligado con el plásmido pGBS19⁸² previamente digerido, desfosforilado y purificado de un gel de agarosa 1%. En el Apéndice I se incluye el artículo "Recovery of DNA from agarose gels stained with Methylene Blue", en donde se describe el protocolo empleado en la purificación del fragmento de DNA correspondiente al vehículo linearizado. El producto de esta reacción de ligasa, fue usado para transformar la cepa de *E. coli* HB101 (*leu*⁻ *pro*⁻). La selección positiva de las clonas portadoras del gene *pac*, estuvo basada en la propiedad de la enzima PA para hidrolizar diferentes compuestos que poseen un grupo fenilo, como la fenil-prolina y la fenil-glicina, de tal manera que la liberación del aminoácido de su correspondiente complejo, permitió complementar las auxotrofías de la cepa.²⁹

Dado que la forma activa de la enzima activa PA consta de las subunidades α y β , y que se ha identificado que la especificidad de la enzima hacia grupos fenilo radica en la subunidad α ,²² mientras que la subunidad β contiene un residuo de serina reactivo¹², se pensó construir un primer conjunto de genes híbridos mediante el intercambio de estas subunidades. Para delimitar y manipular estas dos regiones estructuralmente relevantes, en las regiones de DNA que codifican para los péptidos conectores de las enzimas, se crearon mediante mutagénesis dirigida sitios de restricción de la endonucleasa *Kpn*I. Así mismo, en esta mutagénesis fueron creados los sitios de restricción *Eco*RI y *Hind*III para delimitar a los genes *pac* en sus extremos 5' y 3', respectivamente (Fig. 11). Las mutagénesis realizadas fueron efectuadas mediante un protocolo modificado al propuesto por Melson y Long,⁵⁰ mismas que son reportadas en el artículo "A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC/M13 vectors", incluido en el Apéndice I. Los fragmentos resultantes de la mutagénesis por PCR fueron digeridos con las endonucleasa de restricción *Eco*RI-*Hind*III, y clonados en el polilinker del plásmido pGBS19⁸², dando origen a los plásmidos pPACEC y pPACKC (Fig. 12). A partir de estos plásmidos, y haciendo uso del sitio de restricción de *Kpn*I, fueron purificados los fragmentos de DNA correspondientes a la subunidad α y β de los genes *pac* de *E. coli* y *K. citrophila*. Estos fragmentos adecuadamente combinados y ligados al plásmido pGBS19, dieron origen a los plásmidos p α ECBEC, p α ECBKC, p α KCBKC y p α KCBKC (Fig. 12).

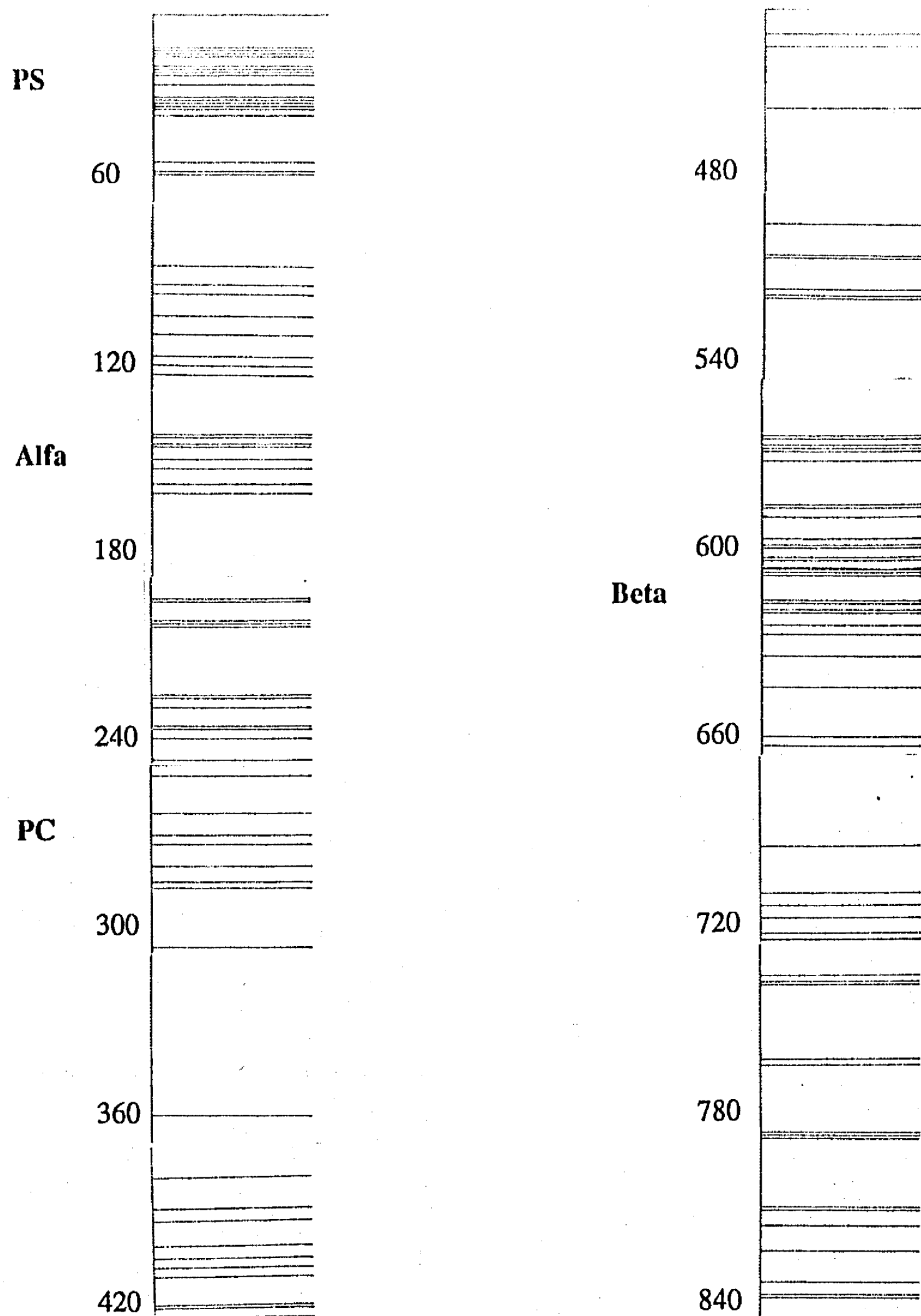


Figura 10. Alineamiento de las secuencias de las enzimas PA de *E. coli* y *K. citrophila*. En la figura se esquematiza el alineamiento mediante barras (aminoácidos diferentes en las proteínas alineadas) y espacios (aminoácidos idénticos en las dos proteínas). PS significa péptido señal, Alfa la subunidad ligera de PA, PC al péptido conector y Beta a la subunidad pesada del precursor de PA.

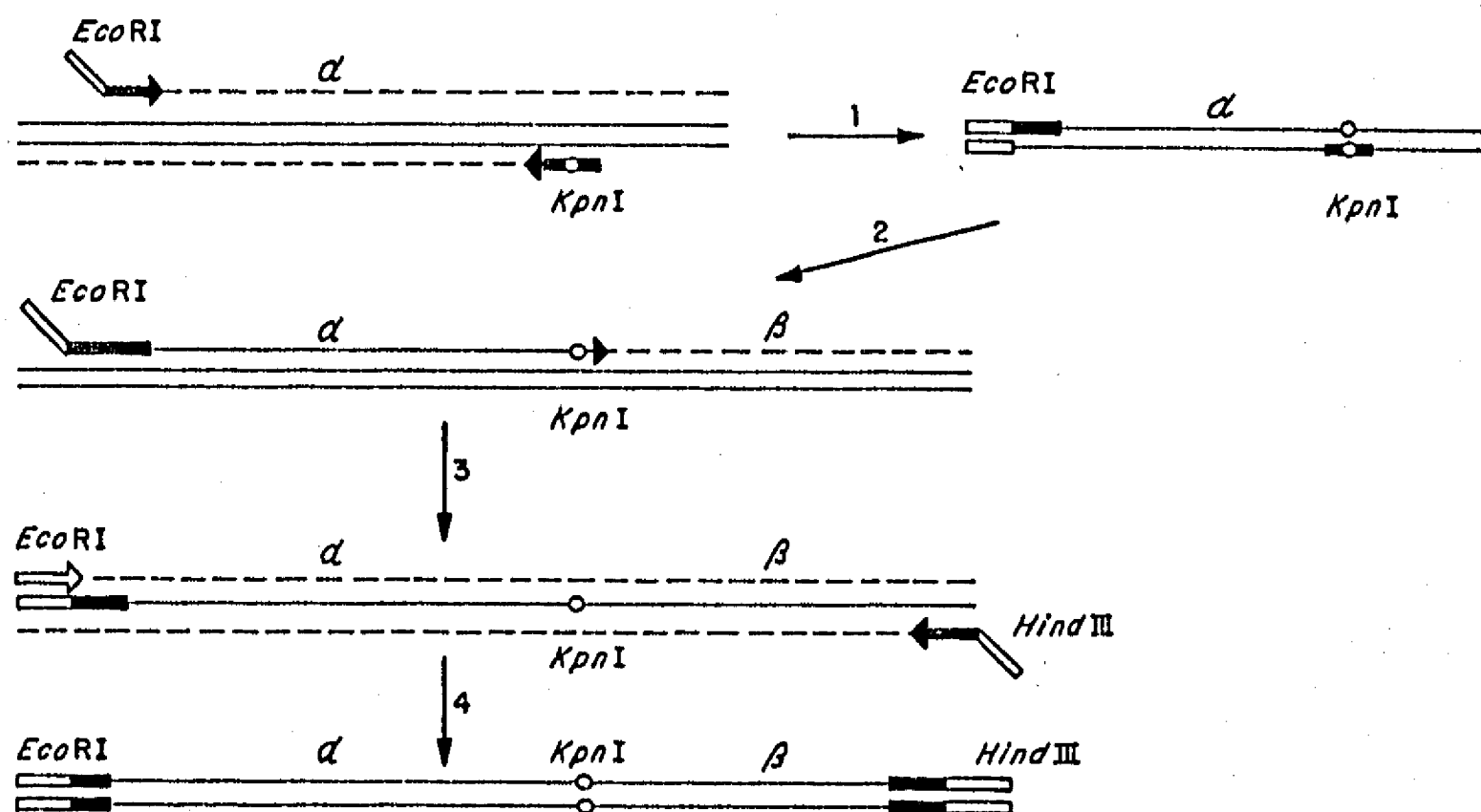


Figura 11. Estrategia de mutagénesis dirigida para la creación de sitios de restricción. En la figura α y β representan a la regiones del gene *pac* que codifican para las subunidades ligera y pesada de la enzima PA, respectivamente. *EcoRI*, *KpnI* y *HindIII* representa a los oligonucleótidos usados en la reacción de amplificación con polimerasa termoestable (PCR). Estos oligonucleótidos introducen en el fragmento de DNA amplificado, los sitios de restricción empleados en la clonación de genes híbridos. Las porciones vacías de los oligonucleótidos, representan regiones de DNA diferentes a las nativas, mientras que las oscuras representan regiones idénticas a las originales.

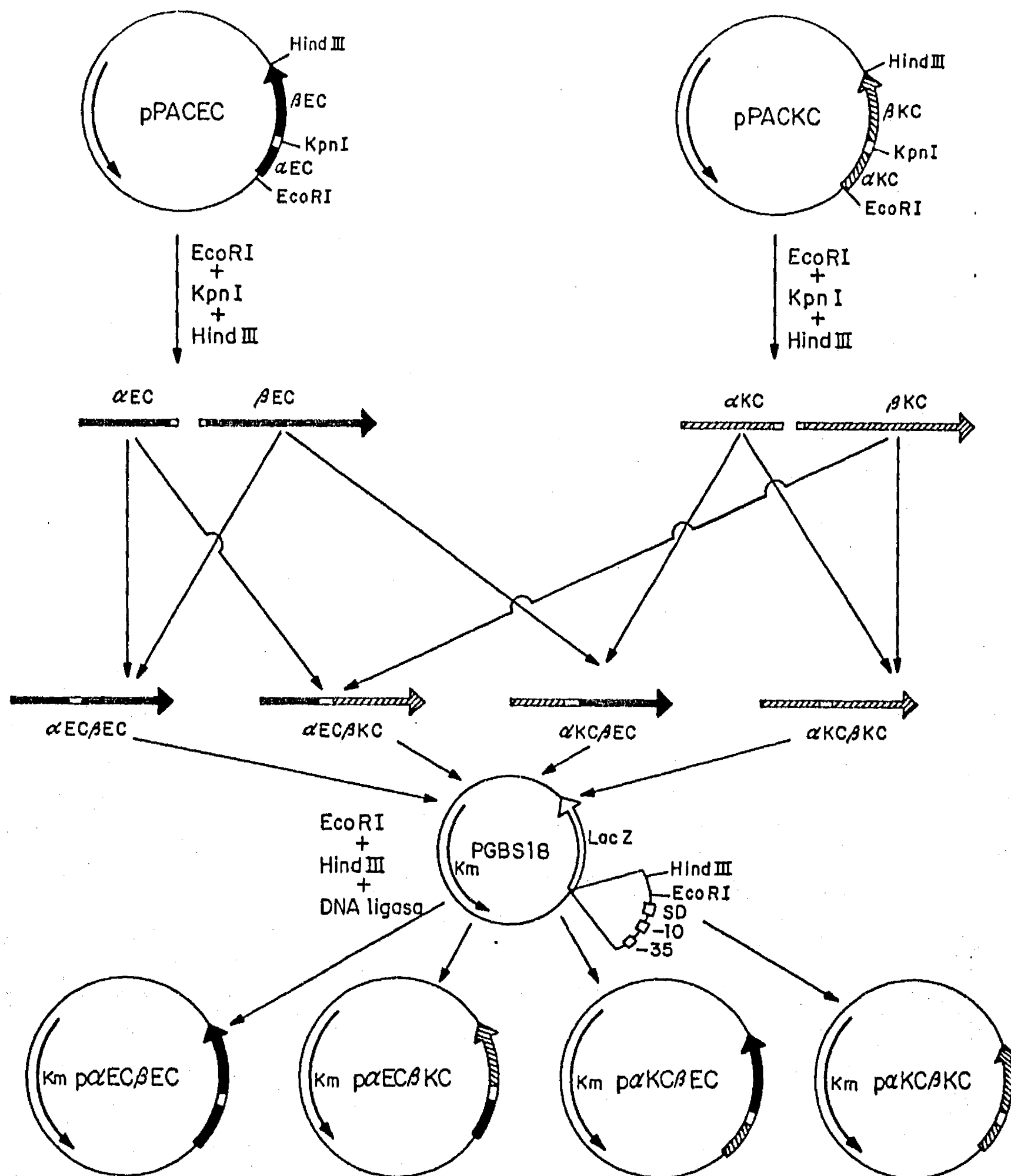


Figura 12. Construcción de genes híbridos de *pac*. En la figura la flecha obscura representa al gene *pac* de *E. coli*, mientras que la flecha rayada al gene *pac* de *K. citrophila*. La región correspondiente al péptido conector se representa por la región abierta de las flechas.

Como puede observarse en la Fig. 10, las proteínas PA de *E. coli* y *K. citrophila*, presentan una gran homología. Aún y cuando las diferencias de residuos entre estas dos proteínas están ubicados a lo largo de las mismas, pueden observarse regiones muy conservadas (*i.e.* residuos 280-380, 98% idénticos) y regiones menos conservadas (*i.e.* residuos 10-35, 40% idénticos, 556-606, 54% idénticos). Es interesante observar que una de las regiones menos conservadas comprende a la región del péptido señal y la parte amino de la subunidad α . El intercambio de lal péptido señal de PA por el de otras proteínas, permitirá elucidar si esta región de la proteína posee un papel en la regulación de la actividad de PA a través del transporte del su precursor.

Es claro que la construcción de proteínas híbridas a partir del intercambio de dominios funcionales permitirá ampliar el horizonte de posibilidades dentro del área de Ingeniería de Proteínas. Actualmente los plásmidos con genes híbridos *pac*, descritos en esta sección, están siendo utilizados para determinar algunas de las propiedades cinéticas y de especificidad a substratos de las proteínas híbridas de PA, así como el patrón de termoregulación de las mismas.

III.3 Conclusiones particulares

- La determinación de la actividad de PA, los estudios de fusión con el gene *lacZ*, y la cuantificación del mRNA del gene *pac*, han demostrado que, la regulación por represión catabólica así como la inducción por AFA del gene *pac*, son llevadas a cabo a nivel transcripcional.
- El análisis estadístico de la región 5' de regulación del gene *pac*, indica que el promotor más probable en esta región corresponde al propuesto por Valle *et al.*⁸⁵
- La termoregulación de la actividad intracelular de PA que se efectúa post-transcripcionalmente. Esta termoregulación no es mediada por inactivación proteolítica a 37 °C, o por procesamiento selectivo a 28 °C.

Capítulo IV.- Incremento de la actividad de penicilino acilasa mediante técnicas de Ingeniería Genética

IV.1. Introducción

Como se mencionó en capítulos anteriores, la enzima PA es de gran importancia en la industria farmacéutica. Con el objeto de obtener una cepa sobreproductora de la actividad de PA, se han realizado una gran variedad de estudios que pueden ser primordialmente agrupados dentro de tres áreas:

a) **Ingeniería bioquímica.** En ésta se encuentran los estudios tendientes a definir las condiciones de crecimiento óptimas para la obtención de cultivos con alta actividad de PA (fuente de carbono, temperatura, aereación, inductores de la actividad, etc.). La cepa de *E. coli* ATCC11105 ha sido usada preferencialmente en procesos comerciales,⁴⁶ siendo algunos de los parámetros utilizados comúnmente: temperatura de crecimiento entre 24 y 30 °C, concentración de oxígeno disuelto menor al 2%, inducción por adición de AFA en una concentración de 0.2% y una fuente de carbono diferente a glucosa.^{35,74}

b) **Genética clásica.** En esta área se han utilizado técnicas de mutagénesis generalizada con agentes químicos o físicos. Dentro de los primeros, los más utilizados han sido el MNNG y el EMS.⁷⁹ Usando estos compuestos, y a partir de cepas que poseen una sola copia del gene *pac* en su cromosoma, se han obtenido organismos sobreproductores de la actividad de PA. En la Tabla IV se compilan algunos ejemplos de este tipo de estudio. Es importante mencionar que el incremento en la actividad de PA reportado en dicha Tabla, es un valor relativo a la actividad inicial de la cepa antes de mutagenizar, mismo que comúnmente no se especifica. Por tal motivo, es difícil evaluar la actividad final de PA de dichas mutantes y en consecuencia es conveniente tomar con reserva tales valores de incremento de actividad de PA.

c) **Ingeniería genética.** La tecnología del rDNA ha hecho posible aislar al gene *pac*, clonarlo en plásmidos multicopia, determinar su secuencia nucleotídica, modificar sus señales de regulación, e incluso sustituirlas por las de otros genes. No obstante, las actividades obtenidas con estas modificaciones no han sido tan altas como las esperadas debido a factores como, inestabilidad del plásmido recombinante, saturación de los mecanismos de transporte y procesamiento del precursor.

La inestabilidad de plásmidos es un fenómeno comúnmente observado en sistemas de alta expresión.^{4,26} En particular para el caso de PA, Deretic *et al.* han reportado que la inestabilidad estructural de plásmidos que portan al gene *pac* es muy alta.²⁶ En cuanto a los sistemas de transporte, se sabe que hay un número determinado de canales de secreción por organismo, y para el caso de *E. coli* se describen cerca de 20,000. Estos canales son compartidos con otros precursores distintos al de PA, por lo que puede existir competencia entre ellos por ser transportados en un sistema saturado^{33,84}.

Existen pocos reportes en donde un esfuerzo conjunto en estas tres áreas haya sido aplicado en forma integral.

Como se mencionó en el capítulo II, un factor importante en la producción de 6-APA, producto de la hidrólisis enzimática de PenG, es el empleo de biocatalizadores enzimáticos. En la actualidad, la mayoría de los biocatalizadores empleados en la producción de 6-APA, utilizan a la enzima PA parcialmente enriquecida e inmovilizada. El costo de este tipo de biocatalizadores es comúnmente muy elevado debido a que su elaboración involucra la purificación de la enzima, sin embargo, la actividad específica de este tipo de biocatalizadores es muy alta. En contraparte, en la literatura existen pocos reportes de biocatalizadores de células inmovilizadas. Su empleo podría ser de gran importancia dado su bajo costo. López-Munguía *et al.*,⁷⁰ han evaluado el costo de producción de 6-APA en términos de la actividad específica del biocatalizador y del número de veces que este puede ser reusado. Del análisis realizado en este estudio, es posible concluir que para elaborar un biocatalizador de células inmovilizadas, es requisito contar con cepas que presenten muy alta actividad específica de PA, mismas que podrían ser obtenidas mediante técnicas de Ingeniería Genética.

Cepa original	Cepa mutagenizada	Agente mutagénico o vector	Incremento en la activ. de PA	Mutante resultante	Ref.
<i>E. coli</i> 9637	<i>E. coli</i> cepa5	Nutricional	120.0	Constitutivo	77
<i>E. coli</i> 9637	<i>E. coli</i> M-1	ND ^a	7.0	ND	16
<i>B. megaterium</i> 1494	<i>B. megaterium</i> KFCC	UV	4.0	Constitutivo	79
<i>E. coli</i>	<i>E. coli</i>	MNNG	2.2	Inducible	42
<i>E. coli</i> IFO13500	<i>E. coli</i> AS2	MNNG	6.0	Inducible	56
<i>B. spaericus</i>	<i>B. subtilis</i>	pOH 38	1.7	Constitutivo	30
<i>E. coli</i>	<i>E. coli</i> KFCC-84-3	ND	2.0	Constitutivo	17
<i>K. citophila</i> 21285	<i>E. coli</i>	ND	30.0	ND	29
<i>B. megaterium</i>	<i>B. megaterium</i>	pPL608	1.2	ND	48, 49
<i>E. coli</i> 194	<i>E. coli</i> pMLV7	pACY184	10.6	Constitutivo	96
<i>E. coli</i> 11105	<i>E. coli</i> 5K/pHM6	pBR322	4.6	Constitutivo	46, 81
<i>E. coli</i> 5K/pHM6	<i>E. coli</i>	UV	1.8	Constitutivo	46, 81
<i>P. rettgeri</i> 31052	<i>P. rettgeri</i> CYC1	Nutricional	150.0	Constitutivo	23, 24
<i>P. rettgeri</i> CYC1	<i>E. coli</i> JW358	pACYC184	1.0	Constitutivo	23, 24

Tabla IV. Cepas obtenidas por mutagénesis y recombinación genética para la sobreproducción de la actividad de PA.⁷⁵

^a Información no disponible.

IV.2. Resultados y Discusión

IV.2.1. Construcción de vehículos moleculares para la sobre-expresión del gene *pac*.

Uno de los primeros requisitos para obtener una cepa con alta actividad de PA, es de que en ella exista una considerable tasa de biosíntesis del precursor de PA. La clonación del gene *pac* en plásmidos multicopia, es sin duda una de las formas más directas para lograrlo. En nuestro laboratorio se ha aislado un fragmento cromosomal de aproximadamente 7.5 Kb que contiene el gene *pac* de *E. coli*, así como sus regiones de control 5' y 3'.⁶⁵ Este fragmento fue clonado en el plásmido pBR339 dando origen al plásmido pPA2. La expresión del gene *pac* en esta primera construcción, esta regulada por el control de las señales nativas de regulación, por lo que se está sujeto a represión catabólica e inducción por AFA. Debido a lo anterior, para lograr una máxima inducción del gene *pac* usando este plásmido, es necesario crecer al organismo transformado en un medio de cultivo con fuentes pobres de carbono y altas concentraciones de AFA, que como se mencionó anteriormente, es un inhibidor del crecimiento celular. Estos dos requisitos son

un impedimento para obtener biomasa alta durante la fermentación, y por tanto se reduce considerablemente el rendimiento del proceso. Con el objeto de contender con este problema, fue construido el plásmido pPA101 a partir de la subclonación del gene *pac* en el polilinker del plásmido pGBS19⁸² (Fig. 13). En este plásmido, la transcripción del gene *pac* se encuentra bajo el control del promotor del gene *lacZ* y puede ser inducido si se utiliza lactosa como fuente de carbono o bien con el inductor gratuito isopropil-tiogalactopiranosido (IPTG). Debido a que en la construcción anterior, la región amino terminal del gene *lacZ* quedó en la misma fase de traducción del gene *pac*, el sitio de restricción *Hind*III, ubicado entre estos dos genes, fue digerido, polimerizado y ligado sobre sí mismo. Estas modificaciones adicionan cuatro pares de bases que evitan la formación del producto de fusión (Fig. 14). El plásmido que tiene estas modificaciones fue designado como pPA102 (Fig. 13).

IV.2.2. Estudio del proceso fermentativo de cepas con actividad de PA.

Con el objeto de caracterizar el sistema de sobre-expresión basado en el plásmido pPA102, este plásmido fue transformado en la cepa de *E. coli* JM101. Las células así transformadas fueron crecidas en fermentador de 2 lt a 28 °C, en un medio Bauer⁷⁶ modificado, que contenía: KH₂PO₄ 3%, Na₂HPO₄ 3%, MgSO₄ 3%, (NH₄)₂HPO₄ 17%. El fermentador fue inoculado con 10% (v/v) de un cultivo crecido en condiciones similares a las antes mencionadas. El pH de la fermentación se mantuvo constante mediante la adición de NaOH. Las condiciones de aereación permitieron mantener al cultivo a una concentración de oxígeno disuelto superior al 15%. La presión de selección de células con plásmido se realizó mediante la adición de Kanamicina (Km), en una concentración de 35µM. La velocidad específica de crecimiento μ , fue pre-seleccionada a un valor de 0.2, y fue modulada por limitación de fuente de carbono. Esta se adicionó de una solución alimentadora que contenía glucosa 40% y extracto de levadura 3%, en base al rendimiento de la cepa para convertir glucosa en biomasa, y tomando en cuenta las siguientes relaciones:

$$\ln(\text{Biomasa}_{\text{FINAL}}/\text{Biomasa}_{\text{INICIAL}}) = \mu \delta t, \quad \text{por lo que: } \text{Biomasa}_{\text{FINAL}} = \text{Biomasa}_{\text{INICIAL}} (e^{\mu \delta t})$$

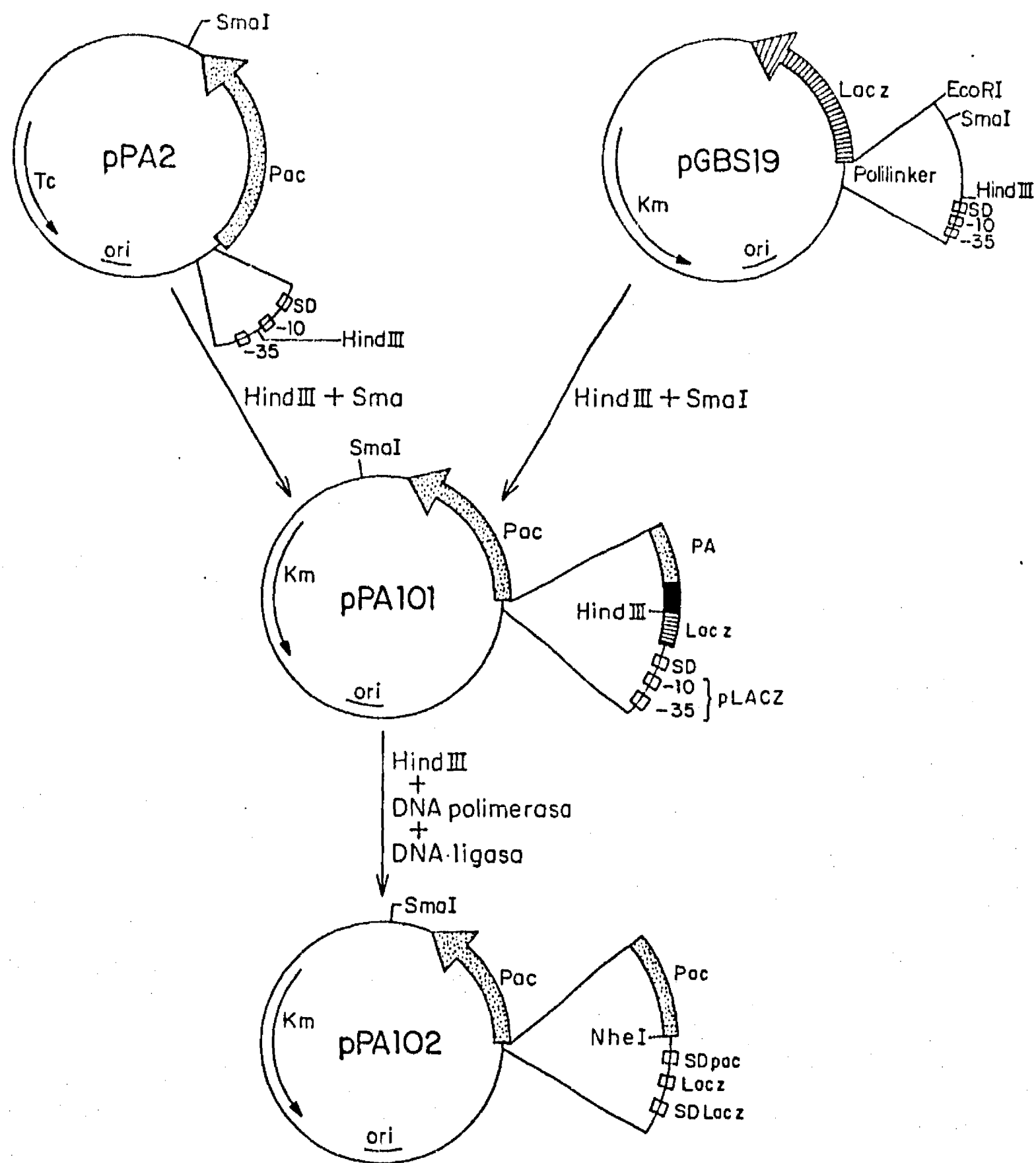


Figura 13. Construcción de vehículos moleculares para la sobre-expresión del gene *pac*. El plásmido pPA101 fue construido al clonar el fragmento *Hind*III-SmaI del plásmido pPA2⁶⁵ en los correspondientes sitios del polilinker del plásmido pGBS19⁸². Con el objeto de evitar la fusión traduccional entre la región amino de *LacZ* y *PA*, el sitio *Hind*III presente en el plásmido pPA101 fue digerido, polimerizado y ligado, dando origen al plásmido pPA102. En este plásmido, la transcripción del gene *pac* se encuentra bajo el control del promotor del gene *lacZ*.

El inductor IPTG se agregó al inicio de la fermentación en una concentración de 100 $\mu\text{M}/\text{ml}$. Las células fueron cosechadas en fase logarítmica tardía, y posteriormente se analizaron los siguientes elementos:

a) La actividad de PA, se determinó usando el método de Balasingham *et al.*,³ descrito en el capítulo III, en donde una unidad de actividad de PA se define como la cantidad de enzima necesaria para formar 1 μmola de 6-APA por minuto a 37 °C. El valor de actividad específica de PA presentó una variación de aproximadamente un 20% entre cada fermentación, y un valor promedio de 0.25.

b) El análisis de proteínas totales se realizó mediante la electroforesis en geles de poliacrilamida al 15%. Este reveló que en las células del cultivo existía una gran acumulación del precursor de PA no procesado equivalente a más del 10% de la proteína celular total (Fig. 15).

c) La estabilidad de plásmido se verificó mediante el análisis de "cuentas viables y por cuantificación directa del "número de plásmidos promedio"⁴¹. Ambos resultados demostraron que la pérdida de plásmido al final de la fermentación fue de aproximadamente un 15%.

d) La biomasa obtenida al final de la fermentación resultó ser de aproximadamente 10 gr/lt peso seco.

IV.2.2.1. Efecto de la concentración y tiempo de adición del inductor del sistema, IPTG, en la actividad de PA.

El análisis de proteínas totales de las fermentaciones anteriores, reveló que gran parte de la enzima PA se acumulaba en forma de precursor, y que por lo tanto los pasos limitantes para la sobre-producción de la enzima activa consistían principalmente en el transporte y/o el procesamiento del precursor.

Dado que la biosíntesis del precursor de PA en estas fermentaciones era mayor al que podría procesarse, y tomando en cuenta de que tales cantidades de precursor podrían tener un efecto negativo, en términos de costo energético celular y saturación de los puntos de transporte a periplasma,^{33,84} se decidió determinar la concentración óptima del inductor para obtener una mayor actividad de PA. Para ello se realizó una fermentación en condiciones similares a las anteriormente descritas, en donde el inductor IPTG fue adicionado a la solución de glucosa y extracto de levadura, de tal manera que la cinética de adición del inductor al fermentador fue de tipo exponencial, tal y como se muestra en la Fig. 16.

La concentración del inductor a la cual se encontró la máxima actividad de PA correspondió a 15-25µM/ml. Es interesante observar que a diferencia de muchos sistemas de sobre-expresión, la condiciones óptimas para una mayor síntesis de la enzima activa, pueden diferir de las condiciones para obtener la mayor actividad del sistema.

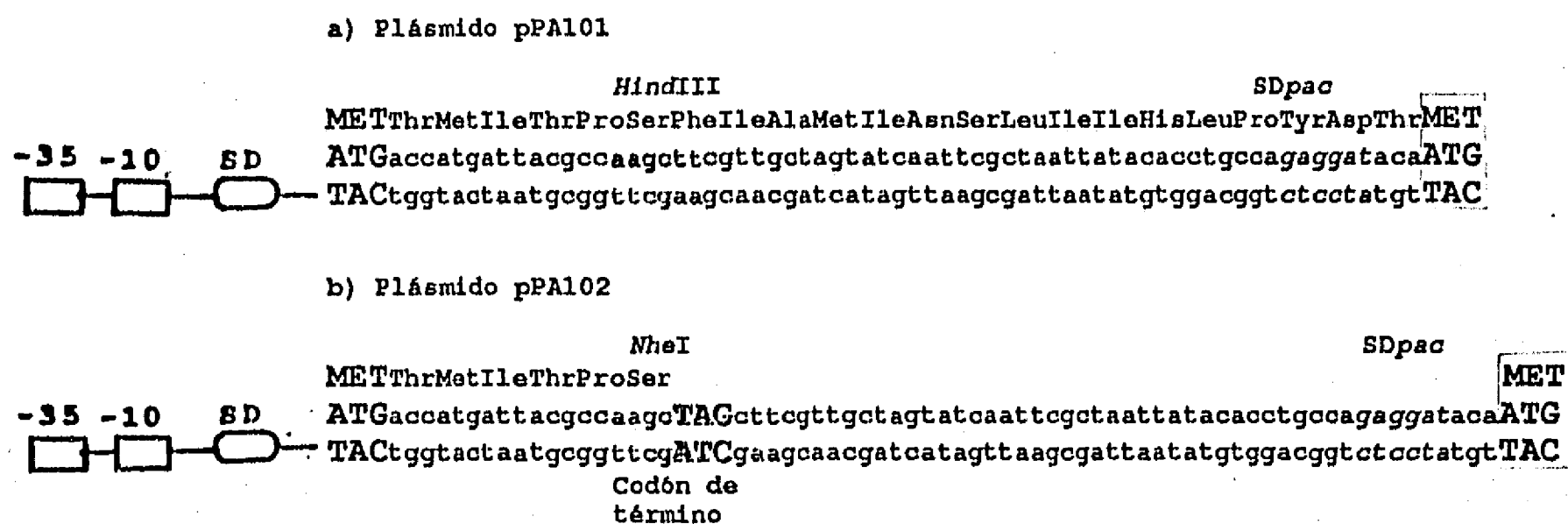


Figura 14. Regiones de regulación involucradas en la expresión del gene *pac* en los plásmidos pPA101 y pPA102. (a) Fusión entre la región 3' del gene *lacZ* y la región de control 5' del gene *pac*. Debido a que estas dos regiones de DNA quedaron en la misma fase de traducción, el producto peptídico proveniente de ellas carece de actividad biológica. Para evitar esta fusión, el sitio de *HindIII* ubicado entre estos dos genes (marcado con letras obscuras), fue digerido, polimerizado y ligado. (b) En el plásmido pPA102 estas modificaciones adicionan cuatro pares de bases con las que se origina un movimiento entre los de los marcos de lectura de estos dos genes. Aunado a lo anterior, con estas modificaciones se creó un sitio de restricción para la enzima *NheI*, que permitió confirmar que dichos cambios fueron producidos correctamente (marcado con letras obscuras). El codón de término TAG, generado con las anteriores modificaciones, se indica con letras mayúsculas. También se muestra el sitio de reconocimiento por el ribosoma, *SD pac*, así como el codón de inicio de la traducción (ATG) del gene *pac*.

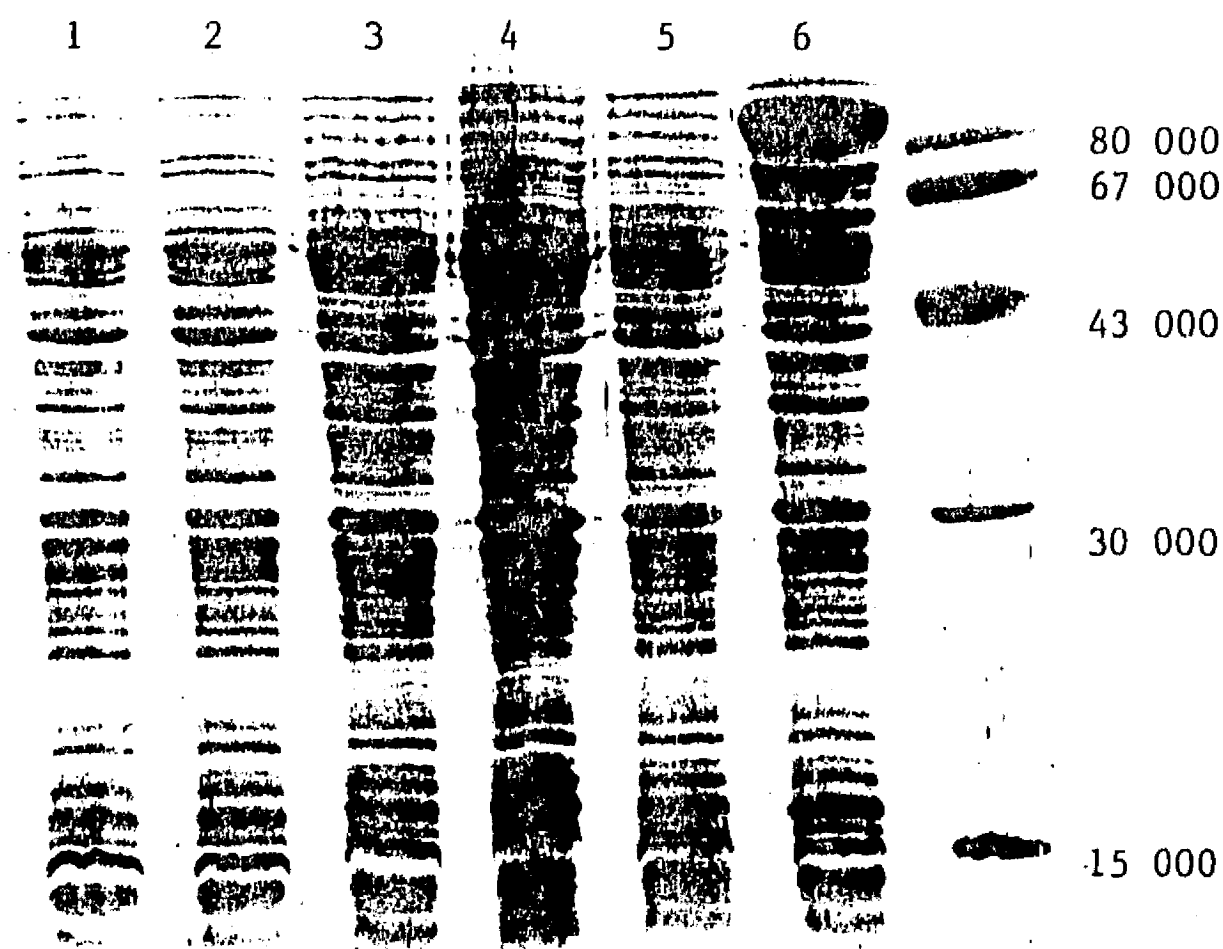


Figura 15. Análisis de proteínas de células sobreproductoras de la actividad de PA. En la figura se muestra el patrón electroforético de geles de acrilamida 15%, de las proteínas totales de las cepas de *E. coli*: 1: JM101, 2: JM101+IPTG, 3: JM101[pGBS19], 4: JM101[pGBS19]+IPTG, 5: JM101[pPA102], 6: JM101[pPA102]+IPTG. El carril de la extrema derecha son marcadores de peso molecular.

Una vez determinada la concentración óptima del IPTG, se realizaron diferentes fermentaciones utilizando las condiciones anteriormente descritas, a excepción de que la concentración del inductor que fue de 25 μ M/ml. En cada una de ellas, el inductor fue adicionado en diferente tiempo. Los perfiles de actividad de PA y crecimiento de estas cepas, se muestran en la Fig. 17.

Como puede observarse en esta figura, la condición de máxima actividad corresponde a tiempos de inducción temprana. En este sentido el sistema de sobre-expresión de actividad de PA, difiere de los sistemas comúnmente empleados, en donde la inducción se efectúa una vez que en el cultivo ha llegado a obtener una alta biomasa.

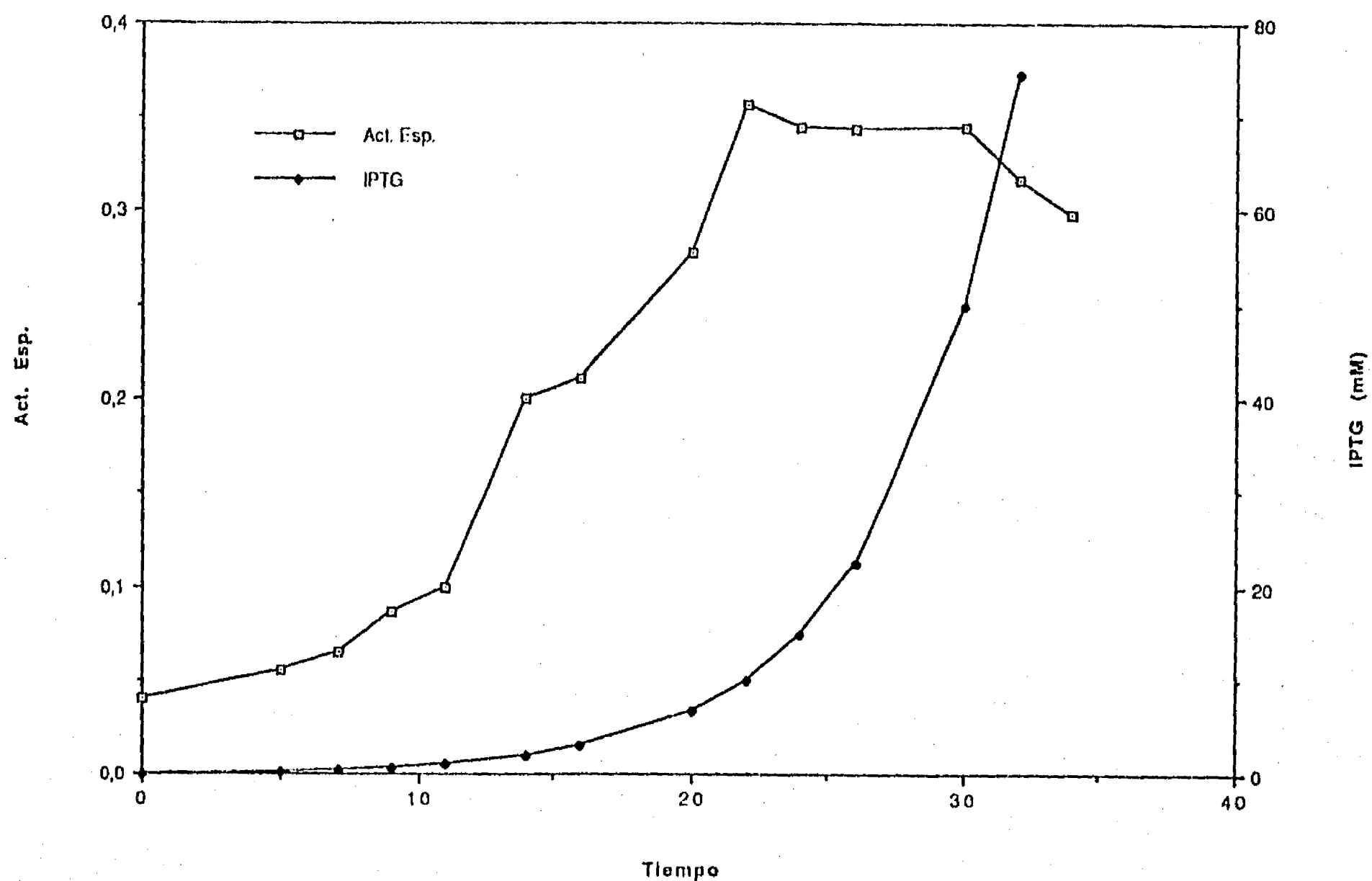


Figura 16. Efecto de la concentración de inductor en la actividad de PA. La cepa de *E. coli* JM101 fue transformada con el plásmido pPA102 y crecida a 28 °C en medio Bauer modificado. La cinética de adición del inductor IPTG al medio, fue de tipo exponencial. En la figura se muestra la concentración final del mismo a lo largo de la fermentación.

Los resultados mostrados en esta sección, son acordes a los obtenidos por Panbangred *et al.*,⁶⁷ en donde se demuestra que la actividad máxima de PA es obtenida al inducir su sistema de expresión bajo la regulación del promotor *tac*, en densidades de cultivo bajas.

IV.2.2.3. Efecto del fondo genético en la actividad de PA.

Es importante considerar que el genotipo de la cepa de *E. coli* ATCC1105, de donde se aisló al gene *pac*, es muy diferente al de las cepas de *E. coli* K-12, comúnmente usadas en el laboratorio, y en particular a la cepa JM101 empleada para realizar los experimentos de fermentación.¹⁰ Aunado a lo anterior, es necesario mencionar que los resultados de los experimentos de tipo Southern presentados en el capítulo III, demuestran que diversas cepas de *E. coli*, entre ellas la cepa JM101, carecen de la región del DNA cromosomal que contiene al gene *pac*. Debido a lo anterior, se pensó que pudieran existir elementos específicos en la cepa de *E. coli* ATCC1105, que favorecieran el transporte y/o el procesamiento del precursor de PA. Por estas razones, la cepa de *E. coli* ATCC1105 fue transformada con el plásmido pPA102, y crecida en fermentador bajo las condiciones empleadas con la cepa JM101[pPA102]. El perfil de actividad de PA y crecimiento celular obtenido en este estudio, es el que se muestra en la Fig. 18.

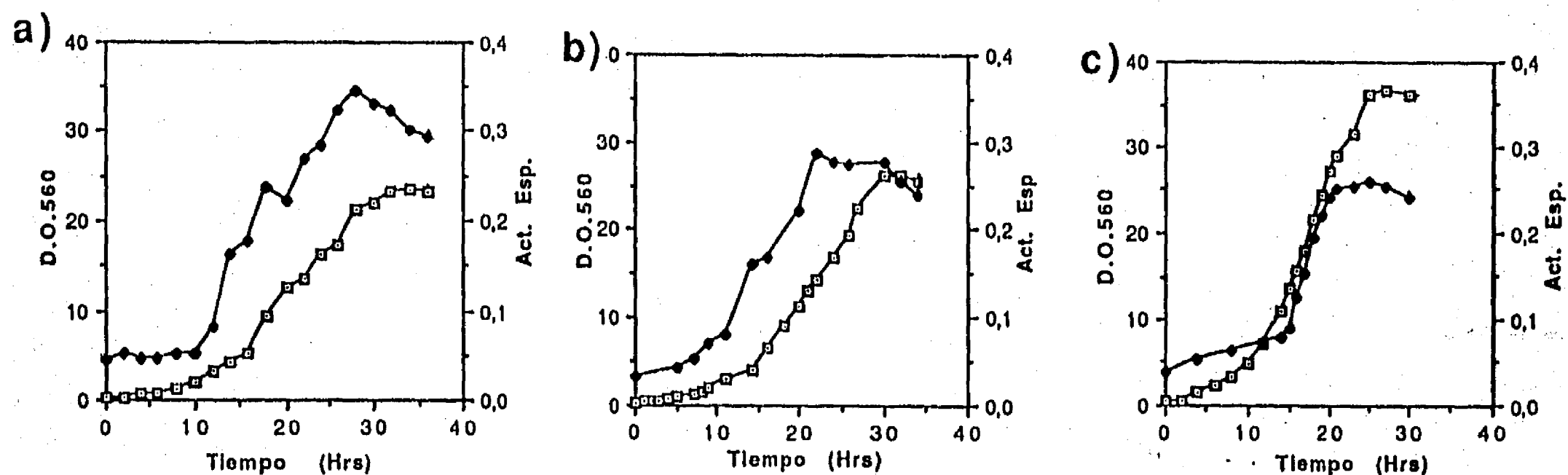


Figura 17. Efecto del tiempo de inducción en la actividad de PA. La cepa de *E. coli* JM101 fue transformada con el plásmido pPA102 y crecida a 28 °C en medio Bauer modificado. La concentración final del inductor IPTG en el medio fue de 25µM/ml. Este fue adicionado en una etapa temprana (a), intermedia (b) o tardía (c), del crecimiento del cultivo celular. —□— D.O.560
—●— Act. Esp.

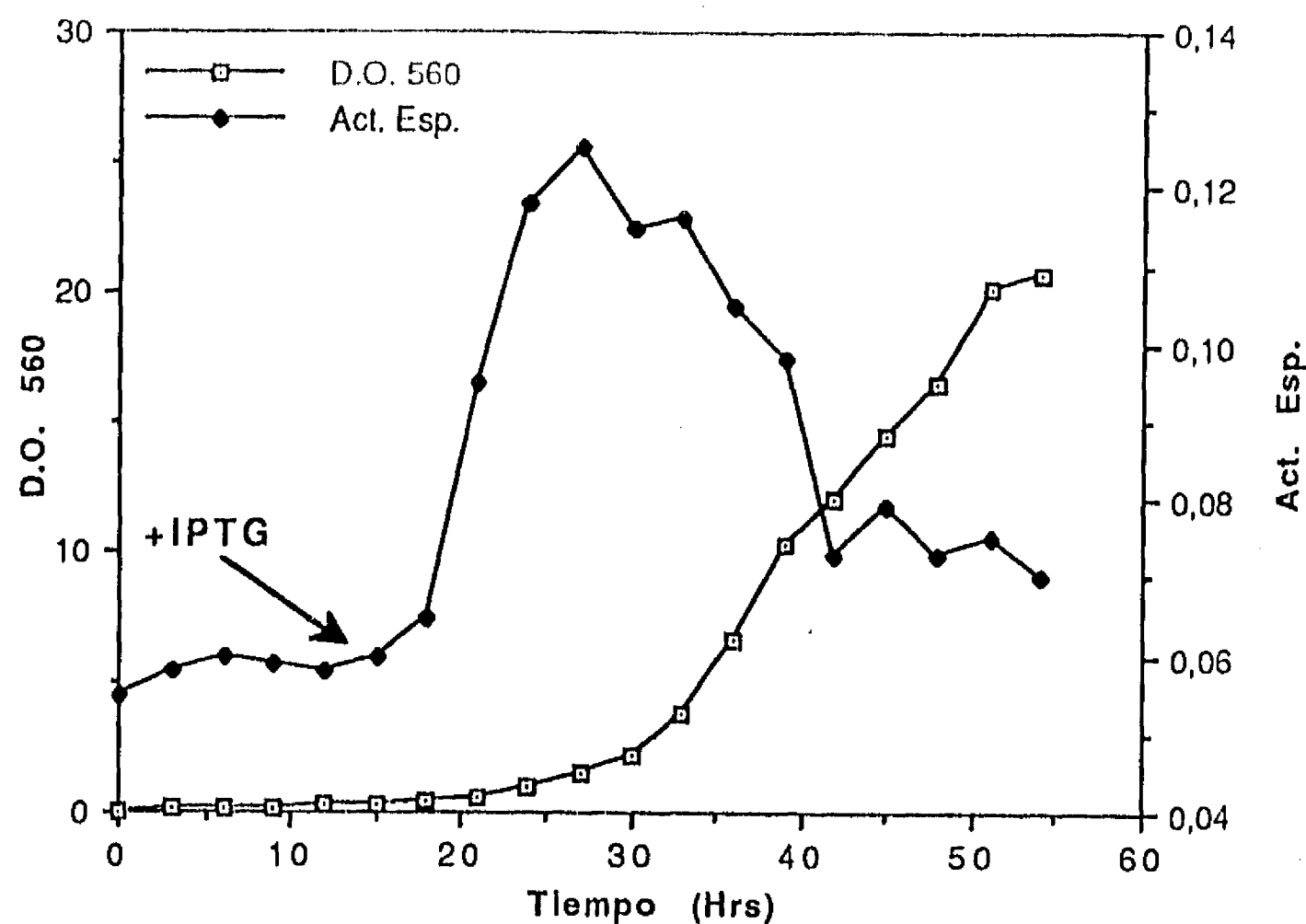


Figura 18. Efecto del fondo genético en la actividad de PA. La cepa de *E. coli* ATCC11105 fue transformada con el plásmido pPA102 y crecida a 28 °C en medio Bauer modificado. El inductor IPTG fue adicionado en una etapa temprana de la fermentación en una concentración final de 25µM/ml.

Como puede observarse en la Fig. 18, la actividad de PA disminuye significativamente a partir de la hora 26. Las determinaciones de Cuentas viables y Número de copias de plásmidos en las células del cultivo, indicaron que no existía una pérdida considerable de plásmidos en la población celular. Cabe mencionar que, pese a que el número de colonias resistentes al antibiótico permanecía constante a lo largo de la fermentación, a partir de la hora 14 y en proporción directa a la pérdida de actividad de PA, se observó que dentro de las colonias estriadas, una población de las mismas presentaba un fenotipo diferente. Este nuevo tipo de colonias no presentó actividad de PA. Los patrones de digestión del plásmido aislado de estas células fue diferente al obtenido del plásmido original pPA102. Estos resultados permiten explicar la pérdida de actividad de PA por inestabilidad estructural del plásmido en la cepa de *E. coli* ATCC11105. Posiblemente tal diferencia de estabilidad en las dos cepas estudiadas sea debida al *locus recA* presente solo en la cepa ATCC11105. El producto peptídico de este *locus* esta involucrado en el proceso de recombinación y posiblemente favoreció rearrreglos en el plásmido pPA102 durante el proceso fermentativo.

IV.3. Conclusiones particulares

- Se desarrolló un proceso para la obtención de cepas de *E. coli* sobreproductoras de la actividad de PA basado en células transformadas con plásmidos que tienen al gene *pac* bajo el promotor de *lacZ*.

- La máxima actividad de PA obtenida en el proceso fermentativo correspondió a cultivos inducidos en una etapa temprana de la fermentación con una concentración final del inductor IPTG de 15-25µM/ml.

- Las modificaciones realizadas para sustituir a las señales nativas de regulación del gene *pac*, por las del gene *lacZ*, permitió obtener un sistema de expresión cuya inducción es independiente del AFA, que es un inhibidor del crecimiento celular. La biomasa obtenida al final del proceso desarrollado en esta Tesis, es aproximadamente tres veces mayor a la obtenida con el primer sistema de expresión.

- La actividad específica del sistema fue en promedio de 0.25 y la producción volumétrica lograda de 1,250 U/lit. De acuerdo al modelo de factibilidad económica de López-Munguía *et al.*,⁷⁰ y considerando la misma estabilidad a la antes obtenida, será necesario triplicar la actividad específica de PA en el proceso obtenido, a fin de poder elaborar un biocatalizador con células inmovilizadas.

- La síntesis del precursor de PA obtenida en el sistema de expresión desarrollado, equivale a más del 10% de la proteína celular total.

- Pese a la alta síntesis del precursor, el incremento de la actividad celular de PA no fue el esperado. El transporte y/o el procesamiento del precursor de PA constituyen los elementos limitantes para el incremento de la actividad de PA en el proceso fermentativo.

Capítulo V. Posible papel de la enzima penicilino acilasa en la Naturaleza

V.1. Introducción

V.1.1. Distribución, propiedades y mecanismo de acción de la enzima PA.

La actividad de PA está presente en una amplia variedad de microorganismos, tales como bacterias (gram positivas y negativas), hongos y levaduras. Ejemplo de microorganismos con esta actividad son entre otros: *Kluyvera citrophila*, *Streptomyces levandulae*, *Proteus rettgeri*, *Actinoplanes sp.*, *Fusarium sp.*, *Bovista plumbea*, *Pseudomonas melanogenum*, *Mycoplana*, *Poraminobacter*, *Aeromonas*, *Xantomonas* y *Acetobacter*.

Las enzimas penicilino acilasas se dividen en tres tipos, de acuerdo a la especificidad de sus substratos y a su fuente de origen⁶⁶:

Tipo I. Son todas aquellas enzimas de hongos y de bacterias que hidrolizan preferencialmente la PenV (fenoximetilpenicilina).

Tipo II. Son enzimas básicamente encontradas en bacterias, y están caracterizadas por su alta especificidad por la PenG (benzilpenicilina).

Tipo III. Son específicas para hidrolizar ampicilina (D-aminobenzilpenicilina) y cefalosporina, como la que poseen *Pseudomonas melanogetum* y *K. citrophila*.

En general las acilasas de tipo II son inhibidas por los productos de la hidrólisis de la PenG, el AFA actúa como un inhibidor competitivo y el 6-APA que inhibe de un modo no competitivo.

La enzima PA de *E. coli* (tipo II), es una de la enzimas más ampliamente usadas en la producción de penicilinas semisintéticas. El 6-APA puede obtenerse también usando PenV, substrato de las PA tipo I.⁵⁵

El mecanismo de acción de la enzima PA es semejante al de las enzimas serin-proteasas. Se ha propuesto que la hidrólisis de PenG por la enzima PA se realiza a través de un intermediario acil-enzima de tal manera que, el grupo acilo donador (i.e. AFA, o un derivado de él) reacciona con la enzima para formar un intermediario que en cambio, transfiere el radical acilo a una amina (i.e. 6-APA), agua o alcohol.³⁶

La enzima no es específica para penicilinas, ya que también hidroliza ciertas cefalosporinas de tipo N y C, aminoácidos, amidas y ésteres. Su especificidad reside en la estructura del grupo acilo y esto la hace diferente de otras enzimas (como las peptidasas), de tal manera que incluso algunos substratos pueden ser hidrolizados más rápidamente que la PenG, tales como: fenilacetamida (110%), N-fenilacetilglicina (182%) y N-fenilacetil-l-alfa aminofenilacético (139%). Un ejemplo de los valores de la constante Km para ciertos substratos hidrolizados por la enzima PA, son los mostrados en la Tabla V.

V.1.2. Vía de utilización del ácido fenilacético en *Pseudomonas* y *E. coli*.

La mayor parte de la información sobre la degradación biológica de compuestos aromáticos proviene de estudios en *Pseudomonas*. Estos organismos son capaces de metabolizar una amplia gama de compuestos aromáticos¹³. No obstante, existen pocos estudios sobre la capacidad de *E. coli* para crecer en compuestos fenilados. Burlingame y Chapman¹³ han demostrado que *E. coli* es capaz de metabolizar, en forma muy específica, ciertos compuestos aromáticos y utilizarlos como fuente de carbono. Algunos de los resultados de dichos autores se presentan en la Tabla VI.

Substrato	Km (M)
Bencilpenicilina (PenG)	$4.6 \cdot 10^{-6}$
Acido 7-aminodesacetoxicefalosporanico	$1.0 \cdot 10^{-5}$
Cefalotina	$4.2 \cdot 10^{-5}$
Etil-fenilacetato	$4.5 \cdot 10^{-5}$
p-nitrofenil-fenilacetato	$3.1 \cdot 10^{-5}$
Fenilacetato p-nitroanilida	$9.7 \cdot 10^{-5}$
Fenilacetilglicina	$8.0 \cdot 10^{-5}$
Cefaloridina	$1.0 \cdot 10^{-4}$
Ampicilina	$5.2 \cdot 10^{-3}$
Acido Bencilpeniciloico	$2.0 \cdot 10^{-3}$
Cefalexina	$2.1 \cdot 10^{-3}$
Acido D-(-)- α -aminofenilacetico p-nitroanilida	$3.2 \cdot 10^{-3}$

Tabla V. Perfil de especificidad de la enzima PA de *E. coli*. Los valores de Km fueron calculados a pH 7.5 a 25 °C⁴³.

Cepa de <i>E. coli</i>	No. de casos	Acidos aromáticos utilizados para crecer					
		AFA	3-HFA	4-HFA	3-FP	3-3HFP	3-HC
B	1	-	+	+	+	+	+
C	1	-	+	+	+	+	+
W	2	+	+	+	+	+	+
K-12	2	+	-	-	+	+	+
NCTC5928	1	+	-	-	+	+	+
Aislados clínicos	19	-	-	-	-	-	-
	5	+	-	-	-	-	-
	2	-	-	-	-	+	+
	8	-	-	-	+	+	+
	1	-	+	+	-	-	-
	2	+	+	+	-	-	-
	3	+	-	-	+	+	+
	27	+	+	+	+	+	+

Tabla VI. Crecimiento de *E. coli* en compuestos fenilados. + Se observó crecimiento.

- No se observó crecimiento. AFA: ácido fenilacético. 3-HFA: 3-hidroxifenilacético.

4-HFA: 4-hidroxifenilacético. 3-FP: 3-fenilpropiónico. 3-3 HFP: 3,3-hidroxifenilpropiónico.

3-HC: ácido-3-hidroxicinámico¹³.

Es interesante comentar que en *E. coli* y *Pseudomonas*, las vías para la utilización del fenil-propiónico, 3-hidroxifenilacético y 4-hidroxifenilacético, están muy conservadas, y por el número de elementos que involucran, son de las vías más complejas descritas en *E. coli*.¹³

Independientemente de la vía para la utilización del fenil-propiónico, algunas cepas de *E. coli* y *Pseudomonas* cuentan con la vía para la asimilación del fenilacético. Esta vía ha sido mapeada cerca del minuto 30 del cromosoma de *E. coli* K-12.¹⁹ Se ha descrito que en esta región del cromosoma se localizan un gran número de genes dispensables para el crecimiento en el laboratorio del organismo, de tal manera que pareciera ser una región silenciosa desde el punto de vista de identificación de mutantes, y por tanto ha sido difícil su caracterización.¹³ Hasta el momento solo se sabe que esta vía es inducible por AFA y que existe una proteína encargada del transporte del mismo.

V.2. Resultados y Discusión

IV.2.1. Crecimiento de cepas de *E. coli* en presencia de compuestos fenilados.

Con base en los resultados de crecimiento de *E. coli* en compuestos fenilados presentados en la Tabla VI, y tomando en cuenta los obtenidos del experimento de "Southern" mostrado en el capítulo III, queda claro que existen cepas de *E. coli* capaces de crecer usando AFA como fuente de carbono, pero que carecen del gene *pac*. En la Tabla antes mencionada, se considera exclusivamente si ciertas cepas de *E. coli* son (+) o no (-) capaces de crecer a expensas de estos compuestos. Con el objeto de observar diferencias cuantitativas en este crecimiento, las cepas de *E. coli* ATCC11105, HB101, JM101 y W3110, fueron crecidas a 28 °C, en medio NN suplementado con AFA 0.1%, como única fuente de carbono (Fig. 19a). Es interesante observar que la cepa que presentó un mayor crecimiento (ATCC11105), es la única que posee al gene *pac*. No obstante, esta diferencia de crecimiento no está asociada directamente a la actividad de PA, ya que se obtienen patrones de crecimiento similares cuando estas cepas son transformadas con el plásmido pMG25 (*pac*⁺) (Fig. 19b). Estos resultados están de acuerdo a los obtenidos en el análisis de la regulación del gene *pac*, y soportan la idea de que este gene pudiera pertenecer a un regulón involucrado en la utilización de compuestos aromáticos.

IV.2.2. Posible papel fisiológico de la enzima PA.

Una de los puntos más importantes del presente proyecto, ha sido el tratar de comprender la función de esta enzima en el organismo. Dado que la enzima PA no confiere resistencia a PenG, queda claro que su función difiere al de las enzimas β -lactamasas, de tal manera que los diferentes nombres que se le da a esta enzima: penicilino acilasa, penicilino amidasa, penicilino amidohidrolasa, etc., solo reflejan el interés del investigador por su aplicación.³⁶ Por otro lado, con base a las propiedades cinéticas de esta enzima, su perfil de especificidad por substrato, anteriormente citado,^{3,43} y considerando el patrón de regulación por fuente de carbono, en el artículo "The role of penicillin amidases in Nature and in industry" (Anexo I), se propone que un posible papel fisiológico de la enzima PA, es el de liberar el grupo fenilo de ciertos compuestos fenilados no metabolizables, para que una vez separado, éste sea utilizado como fuente de carbono.

Con el propósito de verificar la hipótesis anterior, y considerando que la actividad de PA se presenta comúnmente en organismos con actividad de β -lactamasa,³¹ se decidió examinar si una cepa de *E. coli* con estas dos actividades, es capaz de crecer utilizando a la PenG como fuente de carbono. Para ello, la cepa de *E. coli* HB101 fue transformada con el plásmido pMMB22² (β -lactamasa⁺), o simultáneamente con los plásmidos pMMB22 y pPA25⁶⁵ (*pac*⁺). Posteriormente, estas cepas fueron crecidas a 28 °C en medio mínimo NN complementado con casaminoácidos 0.02% y PenG 0.05%, como únicas fuentes de carbono y energía. Las curvas de crecimiento obtenidas de este experimento, se muestran en la Fig. 20. Como puede observarse en dicha figura, el crecimiento de la cepa transformada con el gene *pac*, es al menos del doble con respecto de la que no la tiene, lo cual constituye una evidencia en favor de la hipótesis anteriormente planteada. En el artículo "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene", se analiza de manera más detallada los resultados aquí mostrados (Apéndice I).

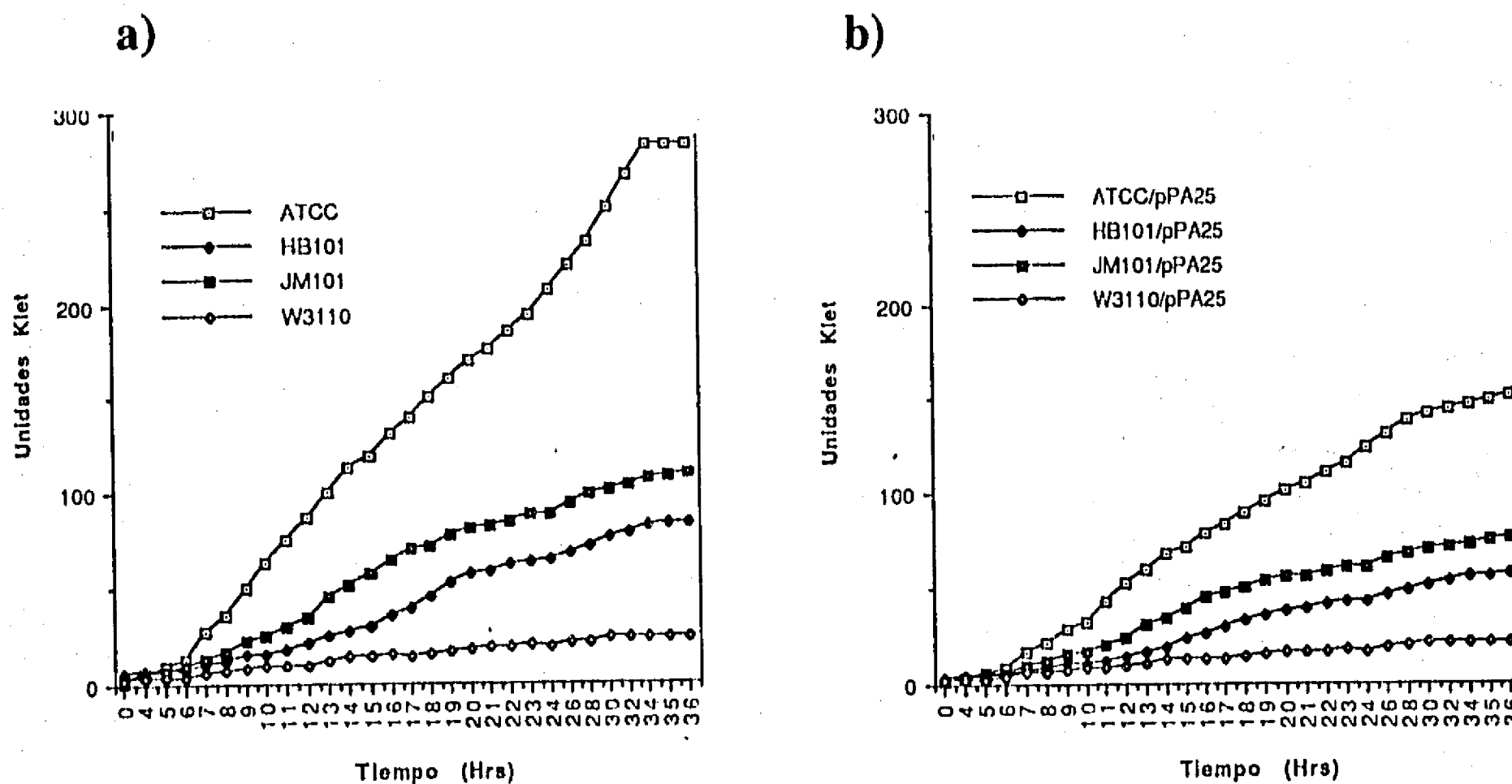


Figura 19. Crecimiento de *E. coli* empleando AFA como fuente de carbono. a) Las cepas de *E. coli* ATCC11105, HB101, JM101 y W3110, fueron crecidas a 28 °C, en medio NN suplementado con AFA 0.1%, como única fuente de carbono. b) Las cepas de *E. coli* ATCC11105[pPA25], HB101pPA25], JM101[pPA25] y W3110[pPA25], crecidas en las mismas condiciones del inciso a.

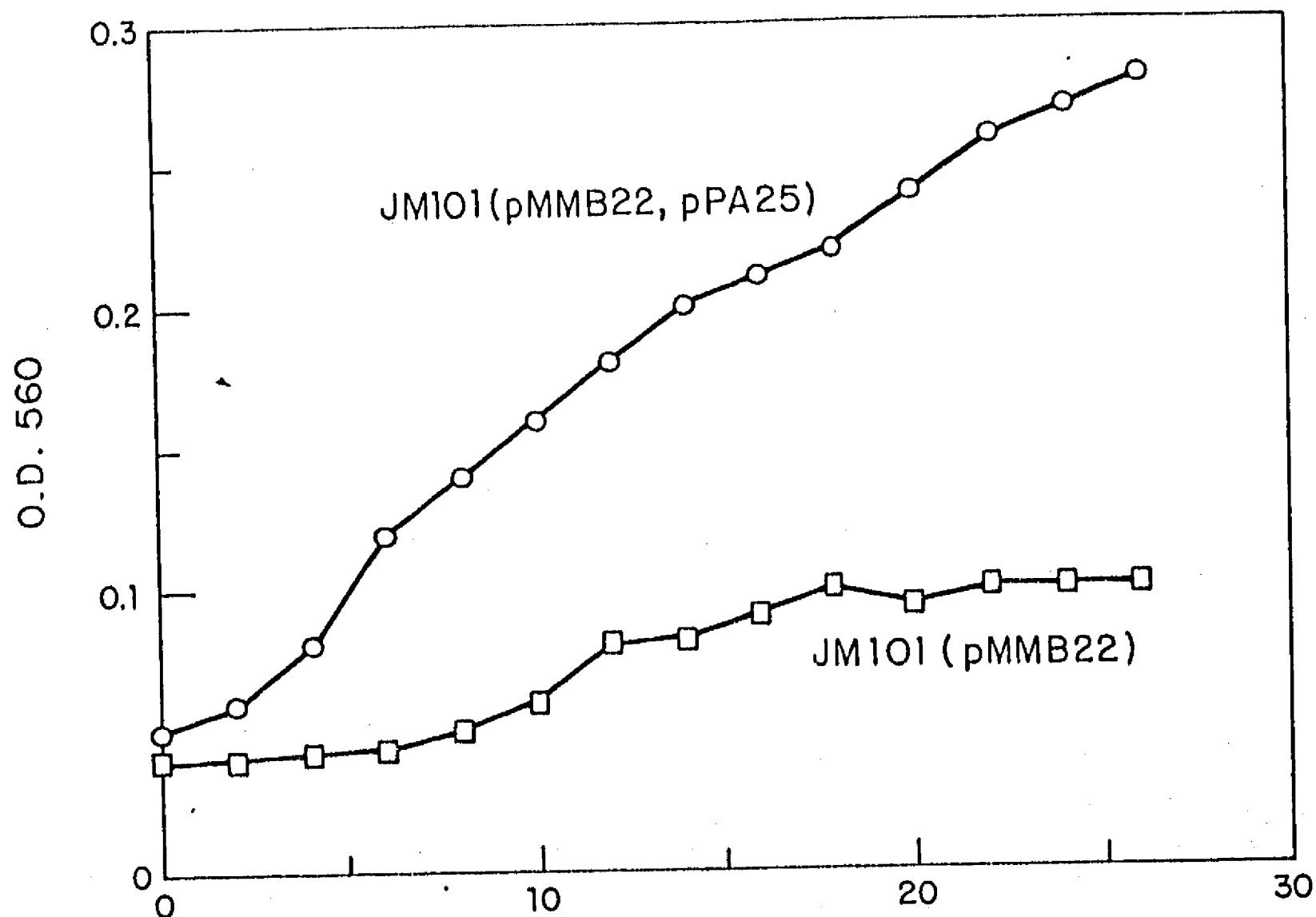


Figura 20. Crecimiento de cepas de *E. coli* empleando PenG como fuente de carbono. Las cepa de *E. coli* HB101[pMMB22] (β -lactamasa⁺) y HB101[pMMB22,pPA25] (β -lactamasa⁺, *pac*⁺) son crecidas a 28 °C en medio mínimo NN, casaminoácidos 0.02% y PenG 0.05% como únicas fuentes de carbono y energía.

V.3. Conclusiones

La actividad de PA puede conferir a cepas de *E. coli* la capacidad de crecer empleando a la PenG como fuente de carbono y energía. En base a este resultado y a los obtenidos en el estudio de regulación de la actividad de PA, y considerandando las propiedades cinéticas y el perfil de especificidad de la enzima PA, se propone que un posible papel fisiológico de esta enzima, es el de liberar el grupo fenilo de ciertos compuestos fenilados no metabolizables, para que una vez separado, éste sea utilizado como fuente de carbono.

Capítulo VI.- Conclusiones generales y perspectivas

Dada la importancia que presenta la enzima penicilino acilasa, tanto en el campo básico, como en el tecnológico, en la presente Tesis se ha tratado de integrar y generar conocimiento relativo a esta enzima utilizando diferentes elementos metodológicos y conceptuales, con un enfoque multidisciplinario.

La determinación de la actividad de PA, los estudios de fusión con el gene *lacZ*, y la cuantificación del mRNA del gene *pac*, han demostrado que, la regulación por represión catabólica así como la inducción por AFA del gene *pac*, son llevadas a cabo a nivel transcripcional. Estos resultados son acordes a los elementos de regulación anteriormente propuestos por nuestro grupo⁸⁵, y junto con el resultado del análisis estadístico empleado en la determinación de posibles promotores⁵⁷, constituyen una evidencia en favor del promotor originalmente propuesto por Valle *et al.*,⁸⁵ como el más pausable.

De manera diferente, la termoregulación de la actividad intracelular de PA se ha demostrado que se efectúa a nivel post-transcripcional aunque aún se desconoce el mecanismo molecular de tal regulación. Los resultados aquí presentados indican que es poco probable que esta termoregulación sea mediada por inactivación proteolítica a 37 °C, o por procesamiento selectivo a 28 °C. Será necesario identificar el evento molecular que determina tal regulación. Mediante experimentos de tipo Western⁸⁰, podría ser identificada la forma (precursor no procesado o subunidades independientes) y la ubicación (periplásmico o citoplásmico) de la enzima PA de células crecidas a diferentes temperaturas. Estos resultados serían clave en la caracterización de este novedoso sistema de regulación.

Como se ha mencionado anteriormente, aún no se han identificado la(s) proteasa(s) involucradas en el proceso de maduración del precursor de PA. Dado que estas proteasas aparentemente están involucradas en el procesamiento de otras proteínas, sería interesante estudiar la relación que pudiera existir entre estas enzimas y la enzima PA. La obtención de mutantes deficientes en el procesamiento del precursor de PA, podría conseguirse mediante mutagénesis química generalizada en cepas transformadas con el gene *pac* en multicopia, de tal manera que se pueda evitar que la pérdida de actividad de PA de dichas mutantes sea por inactivación directa del gene *pac*.

Por otro lado, y con el objetivo de sobreproducir la actividad de PA, se construyeron cepas de *E. coli* en las que se logró un sistema de expresión basado en células transformadas con plásmidos que tienen al gene *pac* bajo el promotor de *lacZ*. Las modificaciones introducidas en la región 5' de control del gene *pac*, permiten que el sistema de expresión sea inducido sin la necesidad del empleo del AFA, que es un inhibidor del crecimiento celular. Por este motivo, la biomasa obtenida al final del cultivo inducido con IPTG resultó ser aproximadamente tres veces mayor a la obtenida empleando AFA como inductor.

Aunado a este incremento en la biomasa, el sistema desarrollado permitió obtener una actividad específica de PA 0.25 y una productividad volumétrica de 1,250 U/lit. Aún y cuando estos valores son superiores a los obtenidos anteriormente, está claro que sería necesario incrementar estos valores en cuando menos tres veces, a fin de poder elaborar de manera rentable, un biocatalizador de células inmovilizadas.

El análisis electroforético de proteínas de las cepas transformadas, demostró que en el cultivo inducido con IPTG, el precursor de PA constituye más del 10% de la proteína celular, y que los pasos limitantes para la sobre-expresión de la actividad de PA, son el transporte y/o el procesamiento del precursor de PA. Aún y cuando el precursor de PA no procesado no incide en la actividad final de PA del sistema fermentativo actual, éste es susceptible de ser convertido en su forma activa mediante futuras implementaciones al sistema de expresión desarrollado. Dichas implementaciones podrían estar basadas en los siguientes enfoques metodológicos:

a) En los estudios de sobre-expresión de la actividad de PA, las únicas cepas de *E. coli* usadas como hospederas del plásmido pPA102, fueron la cepas ATCC11105 y JM101. La búsqueda de nuevas cepas de *E. coli* que pudieran servir como receptoras de este plásmido para el proceso fermentativo, sería una forma sencilla y probablemente eficaz para obtener un mejor sistema de expresión.

b) Como una extensión de la propuesta anterior, podrían ser empleados otros organismos como hospederos del plásmido pPA102 (o derivados del mismo). En particular sería importante utilizar para tal fin a organismos que presentan un eficiente sistema de procesamiento, como son los casos de *Bacillus*, *Streptomyces* o *Aspergillus*.

c) En dado caso que los estudios de regulación demostraran que el transporte del precursor de PA sea un paso limitante en la sobre-expresión de la actividad de PA, sería conveniente realizar, mediante técnicas de Ingeniería Genética, el intercambio del péptido señal de la enzima PA, por el de otras proteínas cuyo transporte sea muy eficiente, como lo es el de la porina OmpA. Este enfoque ha sido empleado exitosamente para secretar otras proteínas como la interleucina 1 β y las enzimas β -lactamasa y subtilisina.⁸⁴

d) Paralelamente a lo mencionado, se ha descrito que las enzimas chaperonas SecB y GroEL ayudan a mantener la conformación de diversos precursores en un estado conformacional compatible con su translocación a través de la membrana.³⁸ Posiblemente la sobre-expresión conjunta de estas chaperonas junto con la de la enzima PA pudiera redituarse en una mayor actividad de PA.

Finalmente, en lo que se refiere al papel de la enzima PA dentro de la fisiología celular, en el presente trabajo se han aportado evidencias que soportan la hipótesis de que el gene *pac* de *E. coli* forma parte de un regulón involucrado en la utilización de compuestos aromáticos, y que particularmente permite liberar a grupos fenilos de ciertos compuestos que originalmente no eran metabolizables. Así mismo, la capacidad que le confiere el gene *pac* al organismo de hidrolizar a la PenG y crecer a partir del AFA liberado, podría ser empleada como elemento de selección en ensayos de actividad de PA. La construcción de cepas derivadas de la *E. coli* ATCC11105 modificadas para obtener el fenotipo *pac*, podrían ser un hospedero adecuado en dicho ensayo.

Capítulo VII,- Otros estudios realizados durante el Doctorado

- Origen y evolución de la transmisión de la información genética

Simultáneamente al proyecto de investigación "Aislamiento y caracterización del gene de la enzima penicilino acilasa (*pac*) de *Escherichia coli*", se realizó un estudio por computadora referente al origen y evolución de los primeros mecanismos de transmisión de la información genética. Los resultados obtenidos en la primera parte del estudio demuestran que, la existencia del patrón RNY (purina- cualquier nucleótido - pirimidina), presente en regiones codificantes del genoma, son independientes del bias originado por la composición de aminoácidos de las proteínas. El resultado encontrado apoya el modelo originalmente propuesta por Crick *et al.*,²⁰ en el que se propone que solo los codones del tipo RNY eran usados por el primer aparato traductor, por lo que era posible identificar la fase correcta del mensaje. Una discusión más amplia de este estudio se presenta en el artículo "New insights on the Comma-less theory" , incluido en el Apéndice I.

- Estudio de genes sobrelapados y sus implicaciones evolutivas

En este estudio se realizó un análisis estadístico de la distribución y tamaño de los marcos abiertos de lectura (ORFs) de la cadena no codificante. El resultado de dicho estudio reveló la existencia de una amplia familia de genes que presentan ORFs sobrelapados en un 100%. Este resultado se interpreta en términos de la posible evolución del sistema información genética de los primeros "organismos vivos" en el manuscrito "Are overlapping open reading frames remnants of a primaeval genetic code system?" incluido en el Apéndice I.

Referencias

- 1 - Aramori I., Fukagawa M., Tsumura M., Iwami M., Ono H., Kojo H., Kohsaka M., Ueda Y., and Imanaka H. (1991) Cloning and nucleotide sequencing of a novel 7 β -(4-Carboxybutamido) cephalosporanic acid acylase gene of *Bacillus laterosporus* and its expression in *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **173**:7848-7855.
- 2 - Bagdasarian, M.M., Amann, E., Lurz, R., Ruckert, B., and Bagdasarian, M. (1983) Activity of the hybrid *trp-lac(tac)* promoter of *Escherichia coli* in *Pseudomonas putida*. Construction of broad-host-range, controlled-expression vectors. *Gene* **26**:273-282.
- 3 - Balasingham D., Warburton P., Dunnill P. and Lilly M. (1972) The isolation and kinetics of penicillin amidase from *E. coli*. *Biochimica et Biophysica Acta.* **8**:250-256.
- 4 - Balbás P., Soberón X., Merino E., Zurita M., Lomeli H., Valle F., Flores N. and Bolivar, F. (1986) Plasmid vector pBR322 and its special-purpose derivatives-a review. *Gene* **50**:3-40.
- 5 - Barbero J.L., Buesa J.M., Gonzáles de Buitrago G., Méndez E, Pérez-Aranda A. and García J.L. (1986) Complete nucleotide sequence of the penicillin acylase gene from *Kluyvera citrophila*. *Gene* **49**:69-80.
- 6 - Batchelor F.R., Doyle F.P., Nayler J.H.C. and Rolinson G.N. (1959) Synthesis of penicillin: 6- Aminopenicillanic acid in penicillin fermentations. *Nature.* **183**:257-258.
- 7 - Bauer S., and Shiloach J. (1974) Maximal exponential growth rate and yield of *E. coli* obtained in Bench-scale Fermentor. *Biotechnology and Bioengineering* **16**:933-941.
- 8 - Böck A., Wirth R. G., Schmid G., Schumscher G., Lang G. and Buckel P. (1983) The penicillin acylase from *Escherichia coli* ATCC11105 consists of two dissimilar subunits. *FEMS Microbiol. Lett.* **20**:135-139.
- 9 - Böck A., Wirth, R., Schmid G., Schumacher G., Lang G. y Buckel, P. (1983) The two subunits of penicillin acylase are processed from a common precursor *FEMS Microbiol. Lett.* **20**:141-144.
- 10 - Brenner D.J., Fanning G.R., Skerman F.J., and Falkow S. (1972) Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. *J. Bacteriol.* **109**:953-965.
- 11 - Bruns W., Hoppe H., Mayer. H.: Penicillin acylase: enzyme and gene structure, a prokaryotic peptide with complex proteolytic processing. 3^{er} European Congress au Biotechnology. Zurich, 10-14 Sep 1984.
- 12 - Bruns W., Hoppe J. and Mayer H. (1985) Structure of the penicillin acylase gene from *Escherichia coli*: A periplasmic enzyme that undergoes multiple proteolytic processing. *Journal of Molecular and Applied Genetics* **3**:36-44.

- 13 - Burlingame R. and Chapman P.J. (1983) Catabolism of the phenylpropionic acid and its 3-hydroxy derivatives by *E. coli*. *J. Bacteriol.* **153**:113-121.
- 14 - Bush K. (1988) β -Lactamase Inhibitors from Laboratory to Clinical. *Microbiology Reviews.* **7**:109-123.
- 15 - Casadaban M.J., Chou J., and Cohen S.N. (1980) *In vitro* gene fusions that join an enzymatically active β -galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals *J. Bacteriol.* **143**:971-980.
- 16.- Cecere F., Olivitri R. and Lavagna S.M. (1985) *ICP* **13**:111.
- 17.- Chang I.S., Cheony K.M. and Soh S. *FRG Patent* 3 437 224/1985.
- 18 - Cole M., Savidge T. and Vanderhaeghe H. (1975) Penicillin acylase (assay). In: Hash J.H. (ed) *Met. enzymol.* **47** 698-721 Academic press.
- 19 - Cooper R.A., Jones D.C., and Parrot S. (1985) Isolation and mapping of *Escherichia coli* K-12 mutants defective in phenylacetate degradation. *J. Gen. Microbiol.* **131**:2753-2757.
- 20 - Crick F.H.C., Griffith J.S., and Orgel L.E. (1957) Codes with out commas. *Proc. Natl. Acad. Sci. USA.* **43**:416-421.
- 21 Daumy G.O., McColl A.S. and Apostolakis, D. (1982) Repression of penicillin G Acylase of *Proteus rettgeri* by Tricarboxylic Acid Cycle Intermediates. *J. Bacteriol.* **152**:104-110.
- 22 - Daumy G.O., and Shaikh K. (1985) Role of Protein Subunits in *Proteus rettgeri* Penicillin G Acylase. *J. Bacteriol.* **163**:1279-1281.
- 23 - Daumy G.O., Danley D., McColl A.S., Apostolakis D., and Vinik F.J. (1985) Experimental evolution of penicillin G acylases from *Escherichia coli* and *Proteus rettgeri*. *J. Bacteriol.* **163**:925.
- 24 - Daumy G.O., Williams J.A., McColl A.S., Zuzel T.J. and Danley, D. (1986) Expression and Regulation of the Penicillin G Acylase gene from *Proteus rettgeri* Cloned in *Escherichia coli*. *J. Bacteriol.* **168**:431-433.
- 25 - Demain L.A. (1985) Industrial microbiology. *Science*, **214**:987-995.
- 26 - Deretic V., Francetic O. and Glisin V. (1984) Instability of the plasmid carrying active penicillin acylase gene from *Escherichia coli*: conditions inducing insertional inactivation. *Microbiology Letters* **24**:173-177.
- 27 - Flores N., Bolivar F., Valle F. and Merino E. (1992) Recovery of DNA from agarose gels stained with Methylene Blue. *BioTechniques*. En prensa.

- 28 - Gang D.M., and Shaikh K. (1976) Regulation of penicillin acylase in *Escherichia coli* by cyclic AMP. *Biochimica et Biophysica Acta*. **45**:110-114.
- 29 - García, J.L., and Buesa, J.M. (1986) An improved method to clone penicillin acylase genes: Cloning and expression in *Escherichia coli* of penicillin G acylase from *Kluyvera citrophila* *Biotechnol. Jour.* **3**:187-195.
- 30.- Gatenbeck S. Nilsson B. Olsson A. and Uhlem M. *International Patent Application* 8 000 929/1986.
- 31 - Hamilton-Miller J.M.T. (1966) Penicillinacylase. *Bacteriol. Rev.* **30**:761-771.
- 32 - Hunt P., Tolley P., Hill P. and Dodson G. (1991) Expression, purification and crystallization of penicillin G acylase from *Escherichia coli* ATCC11105. *Protein Engineering*. **7**:635-639.
- 33 - Ito K., Bassford P.J. and Beckwith J. (1981). Protein localization in *Escherichia coli*: Is there a common step in the secretion of periplasmic and outer membrane proteins. *Cell* **24**:707-717.
- 34 - Kaufmann, W. and Bauer K.: The production of penicillin amidase by *Escherichia coli* ATCC9637. *J. Gen. Microbiol.* **35** (Soc. Gen. Microbiol. Proc. IV), (1964) p.iv.
- 35 - Klein J. and Wagner F. (1980) Immobilization of whole microbial cells for the production of 6-amino penicillanic acid. *Enzyme. Eng.* **5**:335-345.
- 36 - Konecny J. (1981) Penicillin acylases as amidohydrolases and acyl transfer catalysis. *Biotechnol. Letters*. **3**:112-117.
- 37 - Konecny J. (1983) Kinetics of reaction catalyzed by penicillin acylase-type enzymes. In: *Enzyme technology* Berlin, Lafferty R.M. (Ed.): Springer-Verlag. pp. 309-313.
- 38 - Lecker S., Lill R., and Wickner W. (1989) Three pure chaperone proteins of *E. coli* SecB, trigger factor and Gro EL from soluble complexes with precursor proteins *in vitro*. *The EMBO Journal* **8**:2703-2709.
- 39 - Lindsay C.D., and Pain R.H. (1991) Refolding and assembly of penicillin acylase, an enzyme composed of two polypeptide chains that result from a proteolytic activation *Biochemistry*. **30**:9034-9040.
- 40 - Lundrigan, M., and Earhart C.F (1984). Gene *envY* of *Escherichia coli* K-12 affects thermoregulation of major porin expression. *J. Bacteriol.* **157**:262-268.
- 41 - Maniatis T., Fritsch E.F. and Sambrook J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y.
- 42 - Marancbaum E. and Park J.K. (1970) *J. Biotechnol.* **57**:137.

- 43 - Margolin A.L., and Svedas V. (1980) Substrate specificity of penicillin amidase from *E. coli*. *Biochim. Biophys. Acta* **616**:283-289.
- 44 - Matsuda A., and Komatsu K. (1985) Molecular cloning and structure of the gene for 7 β -(4-carboxybutanamido) cephalosporanic acid acylase from *Pseudomonas* strain J. *Bacteriol.* **163**:1222-1228.
- 45 - Matsuda A., Matsuyama K., Yamamoto K., Ichikawa S., and Komatsu K. (1987) Cloning and characterization of the genes for two distinct cephalosporin acylases from a *Pseudomonas* strain J. *Bacteriol.* **169**:5815-5820.
- 46 - Mayer H., Collins J., and Wagner F. (1979) Penicillin G acylase gene of *Escherichia coli* ATCC11105 on multicopy plasmids. In: Plasmid of Medical, Environmental and Commercial Importance, Timmis, K.N. and Phule, A. (eds.), Elsevier/ North-Holland Biomedical Press, New York, pp. 459-470.
- 47 - McCullough E.J. (1983) Gene cloning in *Bacilli* related to enhanced penicillin acylase production. *Biotechnology*. **12**:879-882.
- 48 - McCullough E.J. *UK Patent Application* 2143336/1985.
- 49 - McCullough E.J. *US Patent Application* 509501/1985.
- 50 - Melson R.M. and Long G.L. (1989) A general method of site-specific mutagenesis using a modification of the *Thermus aquaticus* polymerase chain reaction. *Anal. Biochem.* **180**:147-151.
- 51 - Merino E., Valle F., Recillas F., Becerril B., and Bolivar F. (1992) Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene. *Mol. Microbiology* En prensa.
- 52 - Merino E., Osuna J., Bolivar F. and Soberón X. (1992) A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC/M13 vectors. *BioTechniques*. En prensa.
- 53 - Miller H. (1966) Penicillinacylase. *Bacteriol. Rev.* **30**:761-771.
- 54 - Miller J.H. (1976) en Experiments in Molecular Genetics. Cold Spring Harbor Laboratory.
- 55 - Miller R. and Ingolia T. (1989) Cloning and characterization of beta-lactam biosynthetic genes. *Molecular Microbiology*. **5**:689-695.
- 56 - Morita H. and Iwata, T. (1984) Penicillin Acylase Activity in Mutants of *Escherichia coli* Highly Sensitive to Penicillin G. *J. Ferment. Technol.* **62**:217-220.

- 57 - Mulligan, M.E., Hawley, D.K., Entriken, R., and McClure, W.R. (1984) *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acid Res.* 12:789-800.
- 57a - Nakae T. (1986) Outer membrane permeability of bacteria. *Critical Rev. Microbiol.* 13:1-62.
- 58 - Nayler H.J. (1991) Semi-synthetic approaches to novel penicillins. *Trends in Biochemical Sciences* 16:234-237.
- 59 - Neidhardt F.C. and VanBogeken R.A. (1987) Heat shock response. En *Escherichia coli* and *Salmonella typhimurium*. American Society for Microbiology. Washington, D.C. pp. 1334-1344.
- 60 - Oh J., Kim Y., Park Y., Min S. and Kang H. (1987) Complete nucleotide sequence of the penicillin G acylase gene and the flanking regions, and its expression in *E. coli*. *Gene* 56:87-97.
- 61 - Ohashi H., Masao K., Yano M. (1989) Expression of the *Arthrobacter viscosus* Penicillin G Acylase Gene in *E. coli* and *Bacillus subtilis*. *Applied and Environmental Microbiology.* 6:1351-1356.
- 62 - Ohashi H., Katsuda Y., Hashizume T., Abe S.N. Kajiura H., Hattori H., Kamei T., and Yano M. (1988) Molecular cloning of the penicillin acylase gene from *Arthrobacter viscosus* *Applied and Environmental Microbiology.* 54:2603-2607.
- 63 - Okachi R., Misawa M., and Tokashi (1972) Production of 6-aminopenicillanic acid by *Kluyvera citrophila* KY3641. *Agr. Biol. Chem.* 36:925-930.
- 64 - Okachi R., Hashimoto Y., Kawamori R. and Nara T. (1983) Enzymatic Synthesis of penicillins and cephalosporins by penicillin acylase. en *Enzyme Engeenering* (Chibata I., Ed.) vol 6. pp. 81-90.
- 65 - Oliver G., Valle F., Rosetti F., Gómez-Pedorzo M., Santamaría P., Gosset G. y Bolivar F. (1985) A common precursor for the two subunits of the penicillin acylase from *Escherichia coli* ATCC11105. *Gene* 40:9-14.
- 66 - Olivera F., Marjanivic N. and Glisin V. (1988) Molecular biology of penicillin acylase genes. *Biotechnology.* 4:11-15.
- 67 - Panbangred W., Udombunditkul M., and Meevootisom V. (1990) High expression of the penicillin acylase gene in genetically engineered *Escherichia coli*. *Enzyme Engeenering* 10:455-459.
- 68 - Pirt J. (1987) Microbial physiology in the penicillin fermentation. *Trens in Biotech.* 3:64-72.

- 69 - Queener W. and Swartz R. (1989) Penicillins: Biosynthetic and Semisynthetic. *Antimicrobial Agents and Chemtherapy*. 35-95.
- 70 - Rodríguez M.E. (1991) Diseño y caracterización cinética de un biocatalizador de penicilino amidasa. Tesis de Maestría. Instituto de Biotecnología. UNAM.
- 71 - Schumacher, D., Sizmann, D., Haug, H., Buckel, P. and Böck, A. (1986) Penicillin acylase from *E. coli*: unique gene-protein relation. *Nucleic Acids Res.* **14**:5713-5727.
- 72 - Sanders C.C. (1987) Chromosomal cephalosporinases responsible for multiple resistance to newer β -lactam antibiotics. *Ann. Rev. Microbiol.* **41**:573-593.
- 73 - Savidge A.T. (1984) Enzymatic conversions used in the production of penicillins and cephalosporins. *Beechman Pharmaceuticals*. 172-218.
- 74 - Savidge T.A. (1980) Penicillin acylase (Bacterial). *Process Biochem.* **2**:705-721.
- 75 - Shewale G. and Sivaraman H. (1989) Penicillin acylase: Enzyme Production and its application in the manufacture of 6-APA. *Process. Biochemistry*. **7**:146-154.
- 76 - Shimizu M., Okachi R. and Nara T. (1975) Purification and properties of penicillin acylase from *Kluyvera citrophila*. *Agr. Biol. Chem.* **39**:1655-1661.
- 77.- Sikyta B., and Kyslik P. (1981) *Adv. Biotechnol.* **1**:215.
- 78 - Sizmann D., Keilmann C., and Böck A. (1990) Primary structure requirements for the maturation in vivo of penicillin acylase from *Escherichia coli* ATCC 11105. *Eur. J. Biochem.* **192**:192-143.
- 79 - Son J., Miheen T., Seong B. and Han M.H. (1982) Studies on microbial penicillin amidase (IV) the production of penicillin amidase from a partially constitutive mutant of *Bacillus megaterium*. *J.Gen. Appl. Microbiol.* **28**:281-291.
- 80 - Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
- 81 - Stoppok E., Schömer U., Segner A., Mayer H., and Wagner F. (1981) *Adv. Biotechnol.* **3**:547.
- 82 - Spratt G., Hedge P., Heesen S. and Smith B.: Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8 and pEMBL9. *Gene* **4**:337-342 (1986).
- 83 - Szentirmai A. (1964) Production of penicillin acylase *Applied Microbiology* **12**:185-187.
- 84 - Takagi H., Morinaga Y., Tsuchiya M., and Inouye M. (1988) Control of folding of proteins secreted by a high expression secretion vector, pIN-III-ompA: 16-fold increase in production of active subtilisin in *E. coli*. *Biotechnology*. **6**:948-950.

- 85 - Valle F., Gosset G., Tenorio B., Oliver G. and Bolivar F. (1986) Characterization of the regulatory region of the *Escherichia coli* penicillin acylase structural gene. *Gene* 50:119-122.
- 86 - Valle F., Caracterización física y funcional del gene *pac* y de su producto polipeptídico: La enzima penicilino amidohidrolasa. Tesis de Doctorado. U.N.A.M. 1989.
- 87 - Valle F., Balbás P., Merino E., Bolivar F. (1991) The role of penicillin amidases in nature and in industry. *Trends in Biochemical Sciences* 16:36-40.
- 88 - Vandamme J. (1988) Enzymes Involved in β -lactam antibiotic biosynthesis. *Adv. Appl. Microbiol.* 21:89-123.
- 89 - Vandamme E.J. (1981) Penicillin acylase and β -lactamases. In: Economic Microbiology, Rose, A.H. (Ed.), vol 5 Academic press, New York, pp.467-522.
- 90 - Vandamme E.J., and Voets J.P. (1974) Microbial Penicillin Acylases. *Adv. App. Microbiol.* 17:311-369.
- 91 - Vojo. D., Olivera F. and Glisin V. (1984) Instability of the plasmid carrying active penicillin acylase gene from *Escherichia coli* conditions inducing insertional inactivation. *FEMS Microbiol Lett.* 47:173-177.
- 92 - Vojtisek V. and Slezak J. (1976) Penicillinamidohydrolase in *Escherichia coli*. II. Synthesis of the enzyme, Kineties and Speciticity of its induction and the effect of oxygen *Folia Microbiol.* 20:289-297.
- 93 - Vojtisek V. and Slezák, J. (1975) Penicillinamidohydrolase in *Escherichia coli*. III. Catabolite Repression, Diauxie, Effect of cAMP and Nature of the Enzyme Induction. *Folia Microbiol.* 20:298-306.
- 94 - Waxman J., and Strominger L. (1983) Penicillin-Binding proteins and the mechanism of action of β -lactam antibiotics. *Ann. Rev. Biochem.* 52:825-869.
- 95 - Young F.S., and Furano A.V. (1981) Regulation of the synthesis of *E. coli* elongation factor Tu *Cell* 24:695-706.
- 96 - Zhang Q. and Chen L. (1990) Construction and hyperproductivity of engineered strain QE79 bearing recombinant plasmid containing penicillin G acylase gene from *Escherichia coli* strain AS1.76. *Biotechnology Letters* 127:493-498.

Indice de Figuras y Tablas

Figura:

1. Estructura general de un antibiótico β -lactámico.
2. Estructura química de algunas penicilinas semisintéticas.
3. Esquema de la acción de la enzima penicilino acilasa sobre la molécula de penicilina G.
4. Ruta del procesamiento del precursor de PA.
5. Organización física del gene *pac* de *E. coli*.
6. Construcción de vehículos moleculares para el estudio de la transcripción del gene *pac*.
7. Determinación de las actividades de PA, β -galactosidasa y *pac* mRNA.
8. Análisis de los posibles promotores en la región 5' del gene *pac*.
9. Detección del gene *pac* en cepas de *E. coli*.
10. Alineamiento de las secuencias de las enzimas PA de *E. coli* y *K. citrophila*.
11. Estrategia de mutagénesis dirigida para la creación de sitios de restricción.
12. Construcción de genes híbridos de *pac*.
13. Construcción de vehículos moleculares para la sobre-expresión del gene *pac*.
14. Regiones de regulación involucradas en la expresión del gene *pac* en los plásmidos pPA101 y pPA102.
15. Análisis de proteínas de células sobreproductoras de la actividad de PA.
16. Efecto de la concentración de inductor en la actividad de PA.
17. Efecto del tiempo de inducción en la actividad de PA.
18. Efecto del fondo genético en la actividad de PA.
19. Crecimiento de *E. coli* empleando AFA como fuente de carbono.
20. Crecimiento de *E. coli* empleando PenG como fuente de carbono.

Tablas

- I. Propiedades estructurales, cinéticas y regulatorias de enzimas PA.
- II. Condiciones analizadas en el estudio de la regulación transcripcional del gene *pac*.
- III. Estudio del mecanismo de termoregulación de la actividad de PA.
- IV. Cepas obtenidas por mutagénesis y recombinación genética para la sobreproducción de la actividad de PA.
- V. Perfil de especificidad de la enzima PA de *E. coli*.
- VI. Crecimiento de *E. coli* en compuestos fenilados.

Nomenclatura

6-APA	Acido 6-aminopenicilánico.
AFA	Acido fenilacético.
ATCC	American Type Culture Collection.
CRP	Proteína de represión catabólica.
DNA	Acido desoxiribonucleico.
EMS	Etil metano sulfonato.
IPTG	Isopropil-tiogalactopiranosido.
kDa	Kilo Daltons.
Km	Kanamicina.
MNNG	N-metil-N-nitro-N-nitrosoguanidina.
ONPG	Orotidina 5'-monofosfato.
PA	Penicilino acilasa.
pb	Pares de bases.
PCR	Reacción de amplificación con polimerasa termoestable.
PMFS	Fluoruro de fenil-metil-sulfonato.
rDNA	DNA recombinante.
RNA	Acido ribonucleico.
mRNA	RNA mensajero.
UV	Luz ultravioleta.
μ	Velocidad específica de crecimiento.

PERHAPS NO ENZYME has had such an impact in human health as penicillin amidase (PA, EC 3.5.1.11). Its use has made possible a simple procedure for the massive production of a large group of β -lactam semisynthetic antibiotics for the control of bacterial infectious diseases. This enzyme is used to hydrolyse either benzyl-penicillin (penicillin G) or phenoxymethylpenicillin (penicillin V) to produce 6-aminopenicillanic acid (6-APA), the starting compound for the production of semisynthetic penicillins (Fig. 1).

The semisynthetic penicillin derivatives not only exhibit better properties than penicillin G or penicillin V, such as increased stability, easier absorption and fewer side-effects, but represent a practical solution to the problem of adaptive microbial resistance to antibiotics. The most common mechanism of microbial resistance to β -lactam anti-

F. Valle, P. Balbás, E. Merino and F. Bolívar are at the Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Mor. CP. 62271, Mexico.

The role of penicillin amidases in nature and in industry

Fernando Valle, Paulina Balbás, Enrique Merino and Francisco Bolívar

Penicillin amidase (PA) is the enzyme used commercially for the production of semisynthetic penicillins. During the past decade, a detailed picture of the structure and regulation of the gene encoding this enzyme has emerged, revealing a variety of interesting features that are unique among microorganisms. Clues to the biological role of this enzyme have been provided, as well as new strategies for the commercial production and utilization of PA.

biotics is based on a group of enzymes that hydrolyse the β -lactam ring of penicillin, known as β -lactamases. This type of resistance spreads rapidly in bacterial populations by transfer of genetic information among populations and natural selection. Consequently, the massive use of antibiotics soon renders them inefficient, and new antibiotics need to be prepared in order to overcome resistant forms. Cephalosporins, another important group of β -lactam

antibiotics, have been increasingly used in preference to penicillin derivatives. However, the recent introduction of certain β -lactamase inhibitors (clavam compounds) in β -lactam antibiotic preparations has rekindled interest in the use of semisynthetic penicillin derivatives in medicine¹. With its strategic value restored, PA continues to play a key role in the production of useful β -lactam antibiotics for the control of infectious bacterial diseases.

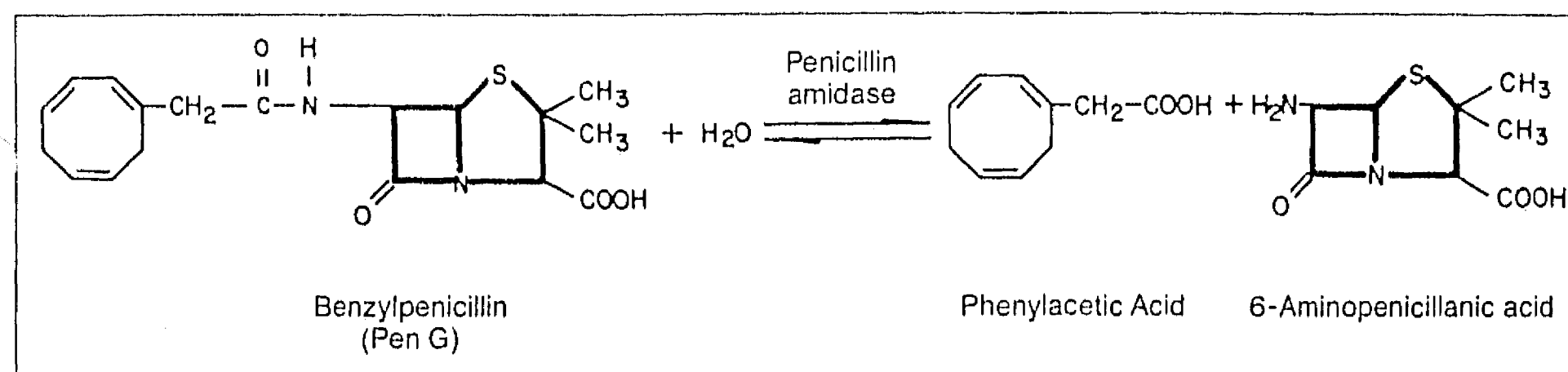


Figure 1

Hydrolysis of penicillin G by PA. The β -lactam ring is indicated with heavy lines.

Penicillin amidase activity is present in a variety of organisms, such as Gram-negative and -positive bacteria, filamentous fungi and yeast. A classification of PAs has been suggested that is based upon the preferential activity of the enzyme over a particular substrate². Type I enzymes preferentially hydrolyse phenoxymethylpenicillin (penicillin V); type II enzymes are primarily active against benzylpenicillin (penicillin G), although they exhibit a broader range of specificities; and finally, type III enzymes (D- α -aminobenzylpenicillin) are most active using ampicillin as a substrate. Industrially, the most commonly used enzyme to date is the type II enzyme isolated from *E. coli* strains ATCC 9637 and ATCC 11105 (for reviews see Refs 2-6).

Penicillin amidase is found in a range of organisms, but its role *in vivo* remains unclear. In fact, the designation of this enzyme as PA reflects its industrial application for the production of 6-APA. A detailed analysis of the gene and protein structures and regulatory mechanisms of cloned and sequenced PA genes has revealed interesting aspects about gene organization, expression and protein processing. This information has led to a hypothesis on the possible biological role of PAs. A better understanding of these elements may be helpful for the design of novel strategies to accomplish overproduction of PA.

An unusual form of protein processing in bacteria

PA from *E. coli* is a periplasmic enzyme that is composed of α - and β -subunits of 23.8 and 62.24 kDa, respectively, as determined by the amino acid sequence inferred from the nucleotide sequence of the gene⁷. However, some heterogeneity in the size of the small subunit becomes apparent⁸ when the enzyme is denatured by SDS-PAGE.

The most striking feature of this enzyme is that the two heterologous subunits that comprise the active molecule are produced from a common polypeptide precursor of approximately 95 kDa. At least three different proteolytic cleavages are necessary to generate the functional enzyme, and none of the processing intermediates exhibits amidohydrolase activity. The first modification of the precursor molecule is the excision of a 26 amino acid leader peptide. This leader permits the transport and translocation of the molecule into the bacterial periplasm⁹. Two subsequent proteolytic cleavages excise a 54 amino acid connecting peptide to produce the separate α - and β -subunits. Amino acid sequence analysis of the purified enzyme has shown that the eight amino acid residues are also removed from the carboxyl terminus of the smallest of the species of the α -subunit⁷. (See Fig. 2 for the linear arrangement of the various fragments.)

Whether the proteolytic modifications take place during the export process or within the periplasmic space is not known⁹. Multistep proteolysis mechanisms exist for many eukaryotic polypeptide hormones, such as insulin, but they are not common in prokaryotic organisms, and thus this complex processing system in *E. coli* is quite noteworthy. The connecting peptides probably serve to direct the correct folding of protein subunits into an active conformation. In the case of PA, this hypothesis is supported by the fact that after denaturation of the active enzyme, regeneration of enzymatic activity is only partially accomplished¹⁰.

Kinetic and biochemical studies have revealed that the highest specificity of the PA enzyme is for the phenylacetyl side chain, and also that the specificity for the penicillin lateral chain is located in the small subunit, while the large subunit possesses a phenylmethanesulfonyl fluoride (PMSF)-sensitive serine

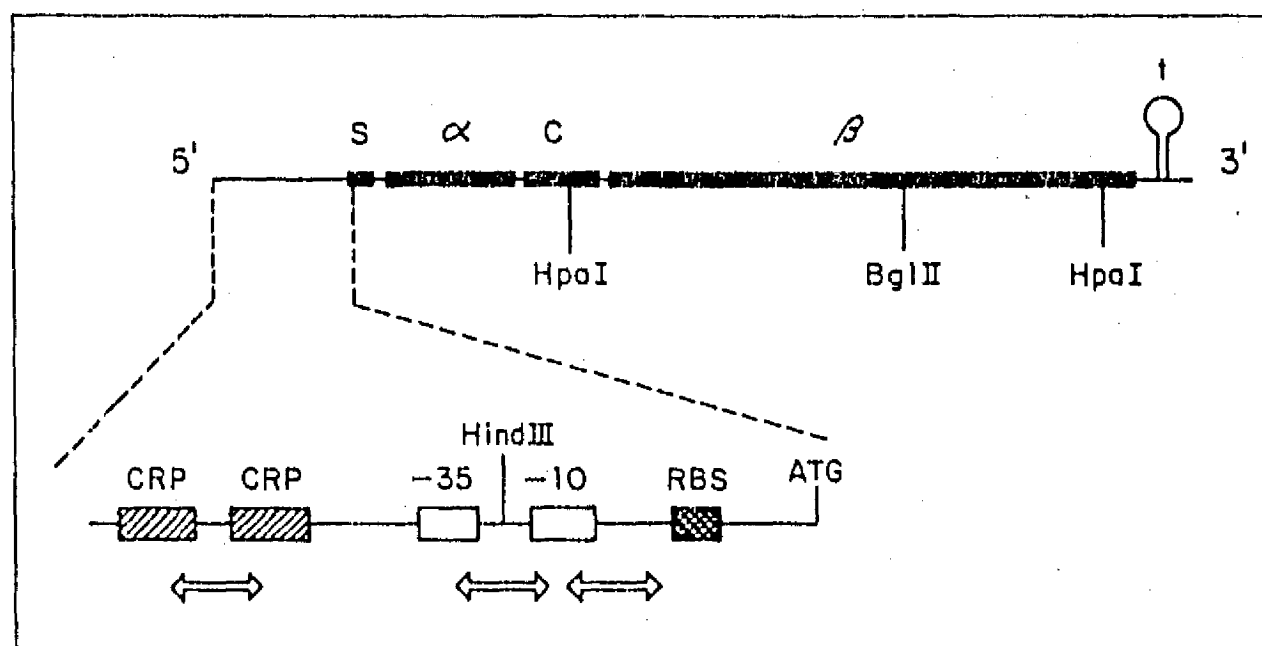


Figure 2

Structure of the *E. coli* PA gene and its 5' and 3' flanking regulatory regions. The 5' regulatory region contains two putative CRP-binding sites. Also indicated are the promoter (−35 and −10 hexamers) imperfect palindromic sequences (open arrows), the ribosome-binding site (RBS), and the translation initiation codon (ATG). The hairpin structure at the end of the PA structural gene (t) is a putative rho-dependent terminator. A few relevant restriction sites are marked and the polypeptide processing units are shown: S, signal peptide; C, connecting peptide; α , small subunit; β , large subunit.

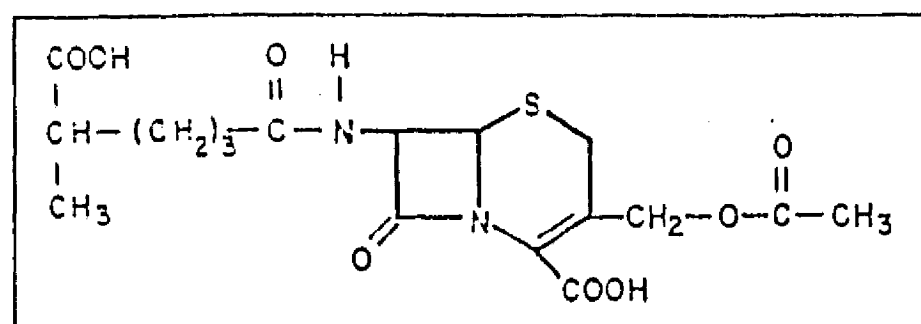


Figure 3

Chemical structure of cephalosporin C. The 7-ACA moiety is emboldened.

residue that plays a key role in the catalytic activity of the enzyme¹⁰. PA is capable of transferring acyl groups from esters, amides or acids to acceptor molecules such as alcohols, amines or water. The reaction mechanism thus resembles that of serine proteases in its ability to form acyl-enzyme intermediate compounds. Phenylacetic acid acts as a competitive inhibitor, while 6-APA is a non-competitive inhibitor. Such a difference is a consequence of the two-step mechanism of the enzyme: the molecules that donate the acyl group compete for the active site of PA, while those that accept the acyl group compete for the acyl-enzyme intermediate¹¹.

A complex regulatory network

There are three major mechanisms regulating PA levels in *E. coli*: cyclic AMP-receptor protein (CRP or CAP), phenylacetic acid¹² and thermoregulation³.

The 5' flanking region of the *E. coli* PA gene contains several DNA sequences that may be involved in the regulation of gene expression (see Fig. 2). Primer extension experiments have suggested that transcription of the PA gene is directed from a single promoter, which exhibits similarity to the *E. coli* σ^{70} consensus promoter¹³. However, promoter fusion experiments have identified another promoter¹⁴. In this promoter, the first ribonucleotide from the mRNA is located two nucleotides before the ribosome-binding site, which is an unusual position in *E. coli* genes. More experiments are required to clarify this point.

Two putative CRP-binding sites have been found around the promoter region, at a position typical of promoters regulated by the CRP mechanism. The 5' regulatory DNA sequence also contains three imperfect palindromic sequences that could be involved in a negative control mechanism¹³.

This array of putative CRP-binding sites, promoter and palindromic

Catabolite repression of the PA gene is observed on a variety of carbon sources such as glycerol, lactose, fructose and glucose. This effect may be partially overcome by the addition of cAMP, which exerts its effect at the transcriptional level. As yet, however, there is no evidence that catabolite repression acts directly at the transcription initiation step, and the observed effect could be mediated by an indirect mechanism, such as the activation of other genes or altered transport of molecules¹². In addition, transcription of the PA gene is induced by the presence of phenylacetic acid as the sole carbon source^{12,15}, thus supporting the hypothesis that PA expression is modulated by carbon sources.

Another interesting feature of PA is its temperature-dependent regulation. Thermoregulation of protein synthesis and protein activity is a poorly studied global regulatory system that is distinct from the heat-shock response¹⁶. In *E. coli*, the optimal temperature for PA activity ranges from 24 to 30°C, and no activity is detectable at 37°C or more^{3,5}. This dramatic effect is not due to heat inactivation of the enzyme, since the optimal temperature for its activity is 45°C (Ref. 3). In *E. coli* there are at least two genetic loci, *entY* and *ompT*, responsible for the thermoregulation of a variety of genes. However, neither of them is involved in the thermoregulation of the PA gene (E. Merino, F. Bolívar and F. Valle, unpublished).

The physiological basis for the thermoregulation of PA and other enzymes in *E. coli* is not understood. Nevertheless, an interesting hypothesis favors the interpretation that *E. coli* recognizes its parasitic or nonparasitic habitats by sensing the temperature; that is, higher temperatures under parasitic growth and lower temperatures in a nonparasitic environment¹⁶. This complex regulatory network would ensure that the PA gene was expressed at low temperatures, with phenylacetic acid as the sole source of carbon.

sequences resembles that of other genes and operons that are subject to carbon regulation. Most of the physiological data available suggest that regulation of the PA gene of *E. coli* is similar to that of genes involved in carbon assimilation.

The PA gene from *Kluyvera citrophila* has been cloned and sequenced and is 87% identical to the *E. coli* enzyme at the amino acid level¹⁷. The PA gene from *K. citrophila*, which is not inducible by phenylacetic acid in this organism, may be induced by this compound when transferred to *E. coli*. This PA gene is also thermoregulated in both organisms, although at different levels¹⁸.

The expression of PA enzymes in *Pseudomonas rettgeri* and *Arrobacter viscosus* is regulated in different ways. In *P. rettgeri* the gene is not inducible by phenylacetic acid and not repressed by glucose, but it is repressed by some tricarboxylic acid cycle intermediates¹⁰, while in *A. viscosus* PA activity is inducible by phenylacetic acid¹⁹.

Is PA from *E. coli* a scavenger for phenylacetylated compounds?

Several groups have addressed the question of the physiological role of PA. The enzyme will recognize preferential enzymatic activity and cleave other phenylacetic acid derivatives in addition to penicillins; some examples of these substrates are presented in Table 1. Together with the regulatory mechanisms of phenylacetic acid induction and catabolite repression, these data suggest that the PA gene is related to pathways involved in the assimilation of aromatic compounds as carbon sources.

Although *E. coli* has been the subject of a variety of biochemical and physiological studies, its ability to grow on aromatic compounds has not been thoroughly investigated. That *E. coli* is able to catabolize such molecules should not be surprising, given the number of aromatic derivatives that are products of the metabolism of a variety of compounds that are present in its intestinal/fecal habitat. Furthermore, an inducible catabolic pathway for the utilization of phenylacetic acid in *E. coli* has been described, and at least two genetic loci have been identified. One of them participates in the transport of phenylacetic acid after its modification to 2-hydroxyphenylacetic acid. However, the cells are incapable of taking 2-hydroxyphenylacetic acid from the environment. It is noteworthy that the genes for this pathway have been mapped to 30.4 minutes on the bacterial chromosome. This region apparently does not carry essential genes for laboratory growth because 60 kDa of DNA from this area may be deleted without generating a different pheno-

type. It has been suggested that this 'silent' region may contain unusual or exotic genes²⁰.

The above evidence suggests that PA is an amidase involved in the degradation of phenylacetylated compounds for the generation of phenylacetic acid, which may be used as a carbon source and could act as an inducer of the degradative pathway. However, this pathway seems to be useless when *E. coli* is living as a parasite, because richer carbon sources are usually in abundance. However, when *E. coli* moves to a free-living state, the utilization of alternative carbon sources becomes obligatory. Phenylacetic acid derivatives are abundant in the nonparasitic environment. For example, they are formed by the action of microflora on plant constituents; these include phenolic and caffeic acids, the flavonoids catechin, myricetin and hesperetin, and the amino acids phenylalanine and tyrosine, as well as some of their metabolites²¹.

The hypothesis that PA is an enzyme involved in the degradation of phenylacetylated compounds is further supported by the thermoregulation of PA gene expression. As mentioned before, when *E. coli* abandons its intestinal habitat, the ability to use the great variety of phenolic compounds available as carbon sources would be an adaptive advantage. Lower temperatures might act as a signal for the detection of the new environment and, in conjunction with the phenylacetylated compounds, the synthesis of PA would then be induced. In the light of this hypothesis, PA should not be an essential enzyme for the parasitic growth of *E. coli*, since at 37°C its enzymatic activity is practically negligible.

Commercial production of PA

A variety of strategies have been designed to increase the commercial production of PA. There have been three major successful approaches to obtain higher PA yields: improvement of the bioengineering process, strain development by conventional microbial genetics, and the use of recombinant DNA methodologies.

The numerous bioengineering approaches have resulted in the optimization of growth conditions for maximal enzymatic activity in large-scale fermentors. For the *E. coli* strain ATCC 11105, which is preferentially used in commercial processes, the optimized parameters for PA production are:

temperature range 24–30°C; dissolved oxygen concentration below 2%; inducer concentration 0.2%; and a carbon source other than glucose^{22,23,24}.

Conventional genetic techniques have been fruitfully exploited to increase the productivity of strains carrying the PA gene. Mutant strains from ATCC 11105 have been isolated on the basis of selection for penicillin G hypersensitivity; the best strains exhibited an eightfold higher activity than the parental strain²⁵. Another approach has been the isolation of mutants (using continuous culture techniques) in the presence of amides as the only source of nitrogen. In addition, glucose has been used in the culture medium to generate mutants that are resistant to catabolite repression; these are fully resistant to repression by glucose, and produce 8–20 times higher PA levels than the parental strain⁵.

The use of genetic engineering techniques has also yielded rewarding results in the development of PA over-producing strains. Recombinant DNA technology has made it possible to alter in a precise, directed manner those specific sequences that regulate gene expression. The most straightforward strategy has been the cloning of the PA genes in multicopy plasmid vectors in order to increase gene dosage from one chromosomal copy to more than 20 (Refs 7,9,15,17). Furthermore, knowledge about the DNA regions, and the molecular mechanisms that modulate PA expression, has made it possible to replace the native regulatory regions for more convenient ones^{7,15}. Although increases of several fold have been obtained in the synthesis of PA precursor, the absolute values in terms of PA activity are not as high as expected. Plasmid instability and saturation of the exporting mechanism are two possible explanations for this result.

Plasmid instability in high-level expression systems is a commonly observed phenomenon. It is affected by several variables in which the host cell genetic background, the induction method, the toxicity of the product and the stability of the expression products

Table 1. Substrate specificities of penicillin amidase from *E. coli*

Substrate	K _m (μM)
Benzylpenicillin (Pen G)	4.6
7-Aminodesacetoxycephalosporanic acid	10
Cephalothin	12
Ethyl phenylacetate	45
p-Nitrophenyl phenylacetate	31
Phenylacetate p-nitroanilide	97
Phenylacetylglutamine	60
Cephaloridine	100
Ampicillin	2200
Benzylpenicilloic acid	2000
Cephalexin	2100
α-(p-nitrophenyl)phenylacetic acid p-nitroanilide	3200

All K_m values were calculated at pH 7.5 and at 25°C²².

(mRNA and protein) play key roles. An extensive review of these issues has been presented elsewhere²⁴. Structural and segregational instabilities have been reported for plasmids carrying the PA genes^{25,26}. To overcome this problem, stable derivatives may be constructed or isolated using rDNA or conventional genetic techniques²⁷.

Saturation of the secretion mechanism may also be a problem in the overproduction of mature PA. The number of secreting channels is thought to be about 20 000 in *E. coli*. If these are shared by several proteins that are secreted into the periplasmic space, transportation of a leader peptide-containing enzyme may become a limiting step. This hypothesis has not been tested experimentally²⁸.

Other strategies for achieving higher PA yields using recombinant DNA have been the use of mutagenesis and strain selection. The mutagenesis schemes have produced mutants with approximately threefold higher activity than the parental strain by modification of the 5' regulatory region, which led to the constitutive synthesis of PA²⁷. The strain selection approach resulted in a more than 50-fold increase in the activity of PA²⁹.

Are cephalosporin-C acetylhydrolases related to penicillin amidases?

As mentioned earlier, cephalosporins are another important group of β-lactam antibiotics. Cephalosporin C, produced by *Cephalosporium acremonium*, is the starting material for most of the currently marketed cephalosporins (Fig. 3). Chemical cleavage produces 7-aminocephalosporanic acid (7-ACA) from which the semisynthetic cephalosporins are derived. However, this process is very delicate and expensive.

Consequently, many pharmaceutical companies and research laboratories have endeavored to develop an enzymatic method for the preparation of semisynthetic cephalosporins analogous to the process developed to produce semisynthetic penicillins. Nonetheless, the search for a commercially useful enzyme has been largely futile.

Some strains carrying cephalosporin-C acetylhydrolases have been reported but none of these activities can efficiently hydrolyse cephalosporin C, and it remains to be seen whether these acetylhydrolases are active on penicillin G or V. The cloning and sequencing of a gene from *Pseudomonas* species, coding for an acetylhydrolase that has activity on 7 β -(4-carboxybutanamido) cephalosporanic acid as a substrate (glutaryl-7ACA acetylhydrolase), revealed some similarity with the *E. coli* PA enzyme. Both enzymes consist of two nonhomologous subunits processed from a common precursor³³, suggesting that they share a common origin^{14,30}.

Another *Pseudomonas* species, strain SE83, produces two distinct glutaryl-7ACA acetylhydrolases¹². Again, both are composed of two different subunits that apparently originated from a common precursor. The structural genes for these two enzymes, *acyI* and *acyII*, have been sequenced³¹. Studies on substrate specificity disclosed a significant difference between them. Nevertheless, the *acyII* enzyme, which has a broad substrate specificity, is capable of hydrolysing cephalosporin C to 7-ACA and aminoadipic acid, although at a very low rate.

There are significant homologies between the amino acid sequences of cephalosporin-C acetylhydrolase II and PA from *E. coli*³⁰. The penicillin amidases from *E. coli*, *K. citrophila*, *A. viscosus*¹⁹, and those enzymes described so far as glutaryl-7ACA acetylhydrolases from *Pseudomonas* strain GK16 and *acyII* from *Pseudomonas* strain

SE83, could share a common origin but have evolved along separate pathways, generating enzymes with different substrate specificity.

Although the cephalosporin-C acetylhydrolases described above do not have an industrially important activity on cephalosporin C as a substrate, they bear a partial resemblance to PA. This similarity with PA is noteworthy, because if PA is really a scavenger for phenylacetylated compounds, then a good cephalosporin-C acetylhydrolase should not be similar to PA, mainly because cephalosporin C does not contain any phenylacetylated groups. Furthermore, this hypothesis suggests that all the cephalosporin-C acetylhydrolases thus far reported, which exhibit very poor activity on cephalosporin C but share similarities with PA, could be modified penicillin amidases that diverged from a common precursor gene. This hypothesis is further supported by the fact that PA from *E. coli* is also capable of hydrolysing certain semisynthetic cephalosporins at a very low rate (Table D). With all these elements in mind, it is possible that protein engineering of penicillin amidases, via site-specific mutagenesis, could be an interesting alternative for the isolation of an efficient cephalosporin-C acetylhydrolase.

Acknowledgements

We are grateful to P. Lizardi for his critical review of the manuscript. This work was partially supported by a grant awarded by Dirección General de Asuntos de Personal Académico, UNAM.

References

- 1 Bush, K. (1955) *Clin. Microbiol. Rev.* 1, 109-123
- 2 Hamilton-Miller, J. M. T. (1966) *Bacteriol. Rev.* 30, 761-771
- 3 Vandamme, E. J. (1981) in *Economic Microbiology Vol. 5* (Rcse, A. H., ed.), pp. 467-522, Academic Press

- 4 Mahajan, P. B. (1984) *Appl. Biochem. Biotechnol.* 9, 537-554
- 5 Shewale, J. G. and Sivaraman, (1989) *Process Biochem.* 24, 146-154
- 6 Francetic, O., Marjanovic, N. and Glisin, V. (1988) *Biotechnologie Apr.* 11-15
- 7 Bruns, W., Hoppe, J., Tsai, H., Bining, H. J., Maywald, F., Collins, J. and Mayer, H. (1985) *J. Mol. Appl. Genet.* 3, 36-44
- 8 Böck, A., Wirth, R., Schmid, G., Schumacher, G., Lang, G. and Buckel, P. (1983) *FEMS Microbiol. Lett.* 20, 141-144
- 9 Schumacher, D., Sizmann, D., Haug, H., Buckel, P. and Böck, A. (1986) *Nucleic Acids Res.* 14, 5713-5727
- 10 Daumy, G. O., Danley, D. and McColl, A. (1955) *J. Bacteriol.* 163, 1279-1281
- 11 Konecny, J. (1981) *Biotechnol. Lett.* 3, 112-117
- 12 Vojtisek, V. and Slezak, J. (1975) *Folia Microbiol.* 20, 298-306
- 13 Valle, F., Gossel, G., Tenorio, B., Oliver, G. and Bolívar, F. (1986) *Gene* 50, 119-122
- 14 Oh, S.-J., Kim, Y.-C., Park, Y.-W., Min, S.-Y., Kim, I.-S. and Kang, H.-S. (1987) *Gene* 56, 87-97
- 15 Oliver, G., Valle, F., Rosetti, F., Gómez-Pedrozo, M., Santamaría, P., Gossel, G. and Bolívar, F. (1985) *Gene* 40, 9-14
- 16 Nakare, T. (1986) *Crit. Rev. Microbiol.* 13, 1-62
- 17 Barbero, J. M., Buesa, J. M., González-de-Suñtrago, G., Méndez, E., Pérez-Aranda, A. and García, J. L. (1986) *Gene* 49, 63-80
- 18 García, J. L. and Buesa, J. M. (1986) *J. Biotechnol.* 3, 187-195
- 19 Ohashi, H., Katsuta, Y., Hashizume, T., Abe, S., Kajura, H., Hattori, H., Kamei, T. and Yano, M. (1988) *Appl. Environ. Microbiol.* 54, 2603-2607
- 20 Cooper, R. A., Jones, D. C. and Farrot, S. (1956) *J. Gen. Microbiol.* 131, 2753-2757
- 21 Burlingame, R. and Chapman, P. J. (1983) *J. Bacteriol.* 155, 113-121
- 22 Klein, J. and Wagner, F. (1980) *Enzyme Eng.* 5, 335-345
- 23 Morita, H. and Iwata, T. (1984) *J. Ferment. Technol.* 62, 217-220
- 24 Balbás, P. and Bolívar, F. (1990) *Meth. Enzymol.* 185, 14-37
- 25 Deretic, V., Francetic, O. and Glisin, V. (1984) *FEMS Microbiol. Lett.* 24, 173-177
- 26 Schömer, U., Segner, A. and Wagner, F. (1984) *Appl. Environ. Microbiol.* 47, 307-312
- 27 Mayer, H., Collins, J. and Wagner, F. (1980) *Enzyme Eng.* 5, 61-69
- 28 Ito, K., Sessford, P. J. and Seckwith, J. (1981) *Cell* 24, 707-717
- 29 Hunt, P. D., Tolley, S. P., Ward, R. J., Hill, C. P. and Dodson, G. G. (1990) *Protein Eng.* 3, 635-639
- 30 Matsuda, A., Toma, K. and Komatsu, K. (1987) *J. Bacteriol.* 169, 5821-5826
- 31 Matsuda, A. and Komatsu, K. (1985) *J. Bacteriol.* 163, 1222-1228
- 32 Margolin, A. L., Svedas, V. and Seregin, I. V. (1980) *Biochim. Biophys. Acta* 616, 283-289

molecular microbiology

Dr. G.K. Schoolnik
Division, Geographic Medicine
Bockman Center 241A
Stanford University School of Medicine
Stanford, CA 94305-5425, USA
(415) 723-8158
(415) 723-1399 FAX

April 16, 1992

Dr. Francisco Bolivar
Departamento de Biologia Molecular
Instituto de Biotecnologia
Universidad Nacional Autonoma de Mexico
APDO
Postal 510-3
Cuernavaca, Morelos 62271, Mexico

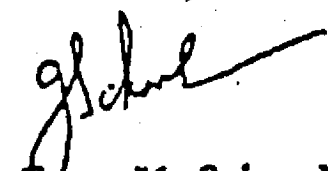
Dear Dr. Bolivar:

Your revised manuscript entitled "Carbon regulation and the role in Nature of the *Escherichia coli* penicillin acylase (*pac*) gene" has been read by Dr. Edmundo Calva and myself. I am pleased to inform you that it has been accepted for publication in Molecular Microbiology.

All future correspondence about this paper should be with Ms. Helen Ure, the Journal's Executive Editor. Her phone number is 011-44-31-226-7232 and her FAX number is 011-44-31-226-3803. Please expect to receive your page proofs in approximately eight weeks.

Thank you for your attention to the Reviewers' comments and for having allowed us to consider this worthwhile contribution.

Sincerely,



Gary K. Schoolnik
Editor

GKS/er
MS EC9203

Editors Dr. C.F. Higgins
Dr. G.K. Schoolnik

Reviews Editor Dr. A.P. Pugsley

Carbon regulation and the role in Nature of the *Escherichia coli*
penicillin acylase (*pac*) gene.

Enrique Merino, Paulina Balbás, Felix Recillas, Baltazar Becerril, Fernando Valle and
Francisco Bolivar.

Departamento de Biología Molecular, Instituto de Biotecnología,
Universidad Nacional Autónoma de México.

Apdo. Postal 510-3, Cuernavaca, Mor. CP 62271, México.

Tel.(5273) 17 23 99; Fax: (5273) 17 23 88.

For correspondence: Dr. Francisco Bolivar.

Running title: Carbon regulation and role in Nature of the *pac* gene.

Summary

Quantitative analysis of specific *pac* mRNA and a *lacZ* fusion to the 5' terminal region of the *pac* gene demonstrated that both, phenylacetic acid (PAA) induction and catabolite repression by glucose are involved, at the transcriptional level, in the regulation of the *pac* gene. The studies presented here suggest that this regulation is also present in *E. coli* transformed strains in which the *pac* gene was not originally present. Analysis of the nucleotide sequence of the 5' terminal region of this gene, with a statistical algorithm, confirms that the putative promoter previously proposed by our group (Valle, *et al.*, 1986), is the most feasible within this region. We demonstrate that the PA activity can confer *E. coli* the ability to use penicillin G (PenG) as a metabolic substrate, by detaching the phenylacetic group which can be used as a carbon source. Based on this data, the regulation properties of the *pac* gene studied in this work and considering the specificity profile of the penicillin acylase enzyme (PA), (Balasingham *et al.*, 1972; Margolin *et al.*, 1980), we suggest its role in *E. coli* as a scavenger enzyme for phenylacetylated compounds.

Introduction

Penicillin acylase (PA) is an enzyme used for the commercial production of semisynthetic penicillins. Several studies have been made to overproduce this enzyme (for a review see Vandame, 1980; Mahajan, 1984; Savidge, 1984; Valle *et al.*, 1991). Despite the fact that PA and its gene have some unique features, many basic aspects on its expression and role in cell metabolism still remain unclear. For example, in *E. coli* ATCC11105, PA is produced as a cytoplasmic precursor that, in order to become catalytically active, has to be transported to the periplasmic space and proteolytically processed (Böck *et al.*, 1983). None of the processing intermediates have PA activity (Schumacher *et al.*, 1986). Interestingly, an internal peptide, located between the two subunits in the PA precursor, is removed in the maturation process to conform the mature enzyme. Another interesting aspect of the enzyme and its gene is that intracellular PA activity is under several regulatory controls, such as temperature, oxygen level, catabolic repression and PAA induction (Kaufman and Bauer, 1964; Szentirmai, 1964; Vojtisek and Slezák 1975a, 1975b). Because of the complex pathway required to render the active form of this enzyme, these regulatory elements may act at different levels, including among others: *pac* transcription, PA transport and PA maturation. Nevertheless, there is no reported data about how this regulation takes place at the molecular level.

Our group has previously published the nucleotide sequence of the 5' regulatory region of the *E. coli pac* gene and established its transcription startpoint 30 bp upstream to the first ATG codon (Oliver *et al.*, 1985, Valle *et al.*, 1986). Analysis of this region has led us to identify two putative CRP-cAMP binding sites, three palindromic sequences that might

be involved in transcriptional regulation, and a previously proposed functional promoter 34 bp upstream of the structural gene (Valle *et al.*, 1986). However, Oh *et al.*, (1987) have proposed a different promoter for this gene.

In this report we extend our analysis of the *pac* gene regulatory region. To study catabolite repression and induction by PAA, we performed quantitative analysis of specific *pac* mRNA and *pac-lacZ* fusion studies. Our results suggest that both PAA induction and catabolite repression by glucose, are involved in the regulation of the *pac* gene at the transcriptional level. Statistical analysis of the 5' terminal region of the *pac* gene confirms that the putative promoter proposed by our group, is the most feasible one in this region.

Finally, the regulatory features of the *pac* gene, as well as the specificity profile of PA, has let us to propose that the role of this enzyme in Nature might be to scavenge for phenylacetylated compounds. In order to test this hypothesis, we performed experiments in which we demonstrate that PA activity can confer *E. coli* the ability to use the phenylacetic group of PenG, as a carbon source, when this bacteria also carries a β -lactamase gene.

Results

PAA induction and glucose repression are mediated at the transcriptional level

To analyze how the *pac* gene is regulated by glucose and PAA, we performed quantitative studies of *pac* mRNA and a *pac-lacZ* fusion. Two different strains were used: *E. coli* ATCC1105, which presents the native PA activity, and *E. coli* HB101 in which the PA activity was not originally present but it was acquired by transformation with plasmid pPA25. In this plasmid, the *pac* gene carries its native 5' regulatory regions (Fig. 1).

Five different conditions were analyzed, and the patterns of PA activity, expression of the PA-LacZ translational fusion, as well as intracellular amounts of *pac* mRNA for tested condition were measured and are depicted in Fig. 2. To make a comparative analysis of these results, the values are expressed in relative terms with respect to the maximum value (100%) of each case. From these results, it can be seen that:

- PA activity is minimal when glucose is used as carbon source.
- An energetically inferior carbon source, such as glycerol, increases PA activity.
- PA activity is maximal when PAA is used as the only carbon source.
- PAA can induce the PA activity even when glucose or glycerol are added to the media, thus indicating that glucose and PAA can simultaneously regulate PA activity.
- These profiles were similar even if the *pac* gene was present in mono or multicopy number.
- In all the above conditions, the PA activities (Fig. 2a) correlate with the amounts of specific *pac* mRNA (Fig. 2d) and PA- β galactosidase determinations (Fig. 2c).

The 5' regulatory elements of the *pac* gene

In order to study the effects of the putative regulatory elements of the *pac* 5' terminal region, we transformed *E. coli* HB101 cells with plasmid pPA4. In this plasmid, the 5' region upstream of the -10 promoter box of the *pac* gene has been removed (Valle *et al.*, 1986, Figs. 1 and 3b). From Fig. 2e, it can be seen that in cells carrying this plasmid, PA activity is no longer affected by glucose repression, PAA does not induce PA activity and a constitutive profile of PA activity is observed in all of the tested conditions.

Computer analysis of the *pac* promoter

In order to identify and evaluate putative σ^{70} promoters within the 5' control region of the *pac* gene, we developed a computer program based on the statistical algorithm proposed by Mulligan *et al.* (1984). Using this program, a maximum value of 60 is obtained when all the bases match the consensus promoter and the -35 and -10 boxes are separated by 17 bp (see *Experimental procedures*). When a cutoff value of 45 was used, the program identified 4 putative promoters within the first 150 bp upstream the structural gene. One of these promoters corresponds to the one previously proposed by our group to be functional, under maximal *pac* induction, and is located 34 bp upstream the first ATG coding codon. It is important to note that using this algorithm, this promoter obtained the maximum score found in the analysis (Fig. 3a). The putative *pac* promoter proposed by Oh *et al.*, (1986) was found when the cutoff value was set down to 37. Using this cutoff value, 14 putative promoters were also found in the analysis.

The *pac* gene is part of a regulon

The results presented in this paper showed that the *pac* gene is regulated in a similar manner in both ATCC11105 and HB101[pPA25] strains. These results indicate that the regulatory elements responsible for PAA induction are present in *E. coli* strains that do not originally have PA activity. This suggests that the *pac* gene shares regulatory elements with other genes present in several *E. coli* strains or that *pac* could be a cryptic gene in some of them. In order to differentiate between these two alternatives, we performed Southern blot hybridization studies using chromosomal DNA isolated from strains that either have (ATCC11105) or do not have (HB101, JM101, W3110) PA activity. The probe used for this experiment was a 1.7 kb *HpaI*-*HpaI* internal fragment of the *pac* gene. As can be seen

in Fig. 4, only the chromosomal DNA from the ATCC11105 strain hybridized, indicating that the *pac* gene is present only in this strain. Furthermore, even when non-stringent washing condition or when a 7.5 kb DNA fragment carrying the entire *pac* gene and chromosomal flanking regions was used as a probe, the same results were obtained (data not shown).

Finally, the size of the *pac* mRNA, determined from the Northern blot studies, was approximately of 3.2 kb (not shown), suggesting that the *pac* gene is transcribed as a mono-cistronic unit.

The PA enzyme confers *E. coli* the ability to grow using the phenylacetic group of PenG, as a carbon source

We have previously proposed that a possible role in Nature of the PA enzyme is to transform non-metabolizable phenylacetylated compounds into ones that can be metabolized (Valle *et al.*, 1991). Taking into account that the coexistence of PA and β -lactamase activities have been observed in several organisms (Hamilton-Miller, 1966), we decided to test if *E. coli* strains with both activities could use PenG as a carbon source. To do so, strain JM101 was transformed with either plasmid pMMB22 (β -lactamase⁺), or with both plasmids, pMMB22 and pPA25, simultaneously (see Fig. 2). These strains were grown in NN minimal media supplemented with casaminoacids and PenG as the only carbon and energy sources. The growth curves obtained are shown in Fig. 5. From this figure it can be seen that the growth rate of cells carrying the *pac* gene was at least twice with respect to the ones without it.

Discussion

It has been previously reported that in *E. coli* the PA activity is under several regulatory controls, including catabolite repression and phenylacetic acid (PAA) induction (Kaufman and Bauer, 1964; Szentirmai, 1964; Voltisek and Sezlák, 1975a, 1975b). Furthermore, it is known that exogenous addition of cAMP partially overcomes catabolite repression by glucose (Gang and Shaikh, 1976) and that maximal PA activity is obtained when PAA is the only carbon source (Voltisek and Sezlak, 1975a). Nevertheless, as far as we know, in all of these studies, it was assumed that the regulation of the *pac* gene directly influenced the intracellular PA activity at the transcriptional level. Although the regulation by glucose of most of the carbon utilization genes is mediated at the transcriptional level, due to the complex pathway needed to get the active form of PA, an extensive analysis was required in order to confirm this. At least there are two ways by which the *pac* gene can be regulated by glucose: a) directly, regulating PA activity by modulating *pac* gene transcription, or b) indirectly, by affecting the expression and/or activity of some other gene(s) involved in the transport or maturation of the PA precursor.

In order to find out which one of these two possibilities was the most plausible, we studied a *pac-lacZ* transcriptional fusion, performed specific *pac* mRNA quantitative analysis and PA activity determinations, in *E. coli* cultures grown under different carbon sources and induction conditions.

Our results indicate that PA and β -galactosidase activities, as well as the amount of specific *pac* mRNA in *E. coli* ATCC11105, are minimal when glucose is used as the only carbon and energy source. This negative effect is partially overcome when a different

carbon source is used, such as glycerol, or when PAA is present in the media. Maximal PA activity was found when PAA was the only carbon source. Interestingly, when the *pac* gene was present in a multicopy plasmid, the same results were observed, suggesting that, at least in the tested conditions, there is no titration effect on a putative regulatory protein (Figs. 2a and 2b). It is also important to mention that the data obtained from the PA- β -galactosidase fusion, and from specific *pac* mRNA studies, are in accordance with these results (Figs. 2c and 2d). However, when most of the original *pac* 5' terminal region was deleted, no regulation effect by glucose or the PAA was observed (Fig. 2e). From these results it can be established that the glucose effect and PAA induction act directly by modulating the transcriptional level of the *pac* gene, and not by an indirect effect on the regulation of other factor(s) involved in the transport or maturation of the PA precursor. Although this was previously assumed, no demonstration of either the transcriptional regulation by glucose or of the effects of PAA induction had been presented before.

A superficial analysis of PA regulation might conduce to inaccurate conclusions. For example, Deretié *et al.* (1984) proposed a correlation between the stability of a plasmid carrying the *pac* gene and the degree of transcriptional activation of *pac*. Nevertheless, the repressed condition used in their study corresponded to a temperature growth of 37 °C, where no PA activity is detected. PA- β galactosidase translational fusion and specific *pac* mRNA quantitations performed in our laboratory, have demonstrated that even when no PA activity is detected, as in cells grown at 37 °C, the *pac* gene is transcribed and PA enzyme is synthesized (data not shown), so there should be another explanation for the stability results obtained by Deretié *et al.*

The results presented in this paper are in agreement with physiological data reported previously (Kaufman and Bauer, 1964; Szentirmai, 1964; Vojtisek and Slezák 1975a, 1975b) and also support the model in which CRP regulates the *pac* gene by interacting with sites located upstream of its promoter (Valle *et al.*, 1986). However, based on the identification of a different promoter, a slightly different model for the *pac* gene regulatory region has been suggested (Oh *et al.*, 1987). This alternative promoter is located six bp downstream of the one proposed by us. Further analysis of this 5' regulatory sequence, using the statistical algorithm developed by Mulligan *et al.* (1984), confirmed that the promoter identified by us is more likely than the one suggested later. Furthermore, the probable mRNA start of the promoter reported by Oh *et al.* is located two nucleotides before the ribosome binding site, which is in an unusual position for *E. coli* genes.

Another interesting result is the fact that similar PA activity profiles were obtained with two different *E. coli* strains: one with a native PA activity (ATCC11105) and another one that did not have the *pac* gene originally (HB101), but was later acquired by transformation. These results indicate that both strains are able to respond in the same way to PAA induction. Two main possibilities may explain the fact that the PAA regulatory elements were present in a strain lacking PA activity. The first one is that the HB101 strain once might have had a functional and regulatable *pac* gene, which later became cryptic. The second possibility is that the regulatory elements responsible for *pac* gene activation by PAA are common to other *loci* of the HB101 strain. To test these hypothesis, we performed a Southern blot hybridization study, in which chromosomal DNA from several *E. coli* strains, able to regulate the *pac* gene present in plasmid pPA25, failed to hybridize when the *pac* gene was used as a probe. These results support the hypothesis of common

regulatory protein(s) involved in PAA regulation. Interestingly, the *Kluyvera citrophila* *pac* gene is regulated in the same fashion when it is transformed into *E. coli* strains, but presents a constitutive expression respect to PAA in *K. citrophila* (García and Buesa, 1986).

Despite many studies on PA and its gene, its role in the cell metabolism is unknown. The results obtained in our studies confirm that the *E. coli* *pac* gene possesses some regulatory features common to many carbon utilization genes and, considering the ability of the PA enzyme to hydrolyze some other phenylacetylated compounds other than PenG (Margolin *et al.*, 1980), we have proposed that the role of this enzyme in Nature is to transform non-metabolizable phenylacetylated compounds into metabolizable ones (Valle *et al.*, 1991). The results presented in Fig. 5, where a higher cell density is reached by a JM101[pMMB22, pPA25] culture, when compared to a JM101[pMMB22] culture, in a minimal media containing PenG as carbon source, are in accordance with this hypothesis.

It is important to mention that PAA is one of the products of PenG hydrolysis by the PA enzyme and that the metabolic degradation of PAA, and aromatic compounds closely related to it (3- and 4-hydroxyphenylacetic acid and phenylpropionic acid), have been previously reported to occur in *E. coli* (Burlingame and Chapman, 1983; Cooper and Skinner, 1980; Cooper *et al.*, 1985). Taking these data into account, we propose that the role of the PA enzyme in *E. coli* is to transform phenylacetylated compounds into metabolizable ones by detaching the phenyl group, which is further used as carbon source. We think that this catabolic role of the *E. coli* PA enzyme may be similar in other organisms, although the metabolic pathway(s) used to assimilate the phenyl group may be different. For example, in *Proteus rettgeri* ATTC31052, PA is not induced by PAA and is

not subject to catabolite repression by glucose, but is repressed by the C₄-dicarboxylic acids of the tricarboxylic acid cycle, such as succinate, fumarate and malate (Daumy *et al.*, 1982). Also, in some fungi, PAA produced by PenG hydrolysis seems to be catabolized to cysteine, acetone and glycine (Shimi *et al.*, 1966). All these data support the broad use of PA enzymes in Nature as scavengers. Further studies shall reveal the relationships of PA enzymes with several catabolic pathways.

Experimental procedures

a) Plasmids

Plasmids pPA4 and pPA25 carrying the *E. coli* *pac* gene have been previously described (Oliver *et al.*, 1985; Valle *et al.*, 1986). Plasmid pMMB22 carrying a β -lactamase gene was constructed by Bagdasarian *et al.* (1983), this plasmid carries an origin of replication compatible with ColE1 replicons.

b) Assays for penicillin acylase and β -galactosidase activities

PA activity was determined by the initial rates of 6-aminopenicilanic acid (6-APA) formation using the colorimetric method described by Balashingham *et al.* (1972). β -galactosidase activity was determined according to Miller (1972). Protein determinations were done as reported by Lowry (1951). Specific activities are expressed in relative terms in order to simplify the analysis.

c) Quantitative analysis of penicillin acylase mRNA

E. coli ATCC11105 and HB101, transformed with plasmid pPA25, were grown to late exponential phase in NN minimal medium supplemented with 15mM NH₄Cl as nitrogen

source. The carbon and energy source used were: a) 0.1% (w/v) of PAA. b) 11 mM glucose c) 22mM glycerol. Total RNA was isolated (Young and Furano, 1981), and loaded onto nitrocellulose membranes using a dot blot camera. Prehybridization was performed at 42 °C for 4 h in a buffer containing 5 X SSC, 50% formamide, 1mM EDTA, 10 mM Tris pH 8, 10 X Denhardt's solution (Denhardt, 1966) and 150 µg/ml of calf thymus DNA. The internal fragment *HpaI-HpaI* of the *pac* gene (Fig. 1) was nick-translated according to Rigby et al. (1977) and used as a ³²P-labeled probe. Hybridization was carried out for 12 h at 42 °C using the prehybridization buffer mixed with the probe. Filters were subsequently washed for 10 min with: 2X SSC, 0.1% SDS at room temperature; 0.1 X SSC, 0.1% SDS at 48 °C; and 0.1X SSC at 64 °C. The radioactivity in each dot was determined by scintillation counting.

d) Northern blot hybridization

The mRNA isolated in the dot blot analysis was electrophoresed on a 1% agarose-2.2M formaldehyde gel for 10 h at 30 V (Lehrach *et al.*, 1977) and then transferred onto nitrocellulose membranes (Southern, 1975). The prehybridization, hybridization and washing procedures were the same as the ones used in the dot blot analysis. In order to determine the position of the 23S and 16S rRNA, filters were stained with methylene blue.

e) Southern blot hybridization

Total DNA from *E. coli* ATCC11105, HB101, JM101 and W3110 strains was extracted and digested with *EcoRI* enzyme. The resulting fragments were resolved by electrophoresis through a 1% agarose gel and blotted onto nitrocellulose membranes (Southern, 1975). The subsequent steps of prehybridization, hybridization and washing procedures were the same as the ones used in the Northern study.

f) Growth of *E. coli* with of PenG as carbon source

E. coli JM101 strain was transformed with plasmid pMMB22 or cotransformed with plasmids pMMB22 and pPA25. The transformed strains were grown at 28 °C on NN minimal media supplemented with 0.02% casaminoacids and 0.05% PenG as sole carbon and energy sources.

g) Computer program to identify σ^{70} promoters

The program developed to search for σ^{70} promoters is based on the statistical algorithm proposed by Mulligan *et al.* (1984) and was written in Turbo C language for IBMpc computers. In this algorithm, each base of the -35 and -10 boxes as well as the ones nearby them, are weighted accordingly to a score matrix. This score matrix was built up based on the frequency of occurrence of each base on a list of *E. coli* and phages σ^{70} promoters. The space between the -35 and -10 regions contributes in geometrical terms to the final score. A cutoff value is used to select only the best promoters within a DNA region.

Acknowledgments

We thank Gloria Soberón for fruitful and gratifying discussions and Noemí Flores and Mercedes Enzaldo for technical support. This work was partially supported by a grant awarded by the Dirección General de Asuntos del Personal Académico, UNAM.

References

- Bagdasarian, M.M., Amann, E., Lurz, R., Ruckert, B., and Bagdasarian, M. (1983) Activity of the hybrid *trp-lac(tac)* promoter of *Escherichia coli* in *Pseudomonas putida*. Construction of broad-host-range, controlled-expression vectors. *Gene* 26:273-282.
- Balasingham, K., Warburton, P., Dunnill, P., and Lilly, M.D. (1972) The isolation and kinetics of penicillin amidase from *Escherichia coli*. *Biochim. Biophys. Acta*. 276:250-256.
- Böck, A., Wirth, G., Schmid, G., Schumacher, G., and Buckel, P. (1983) The two subunits of penicillin acylase from a common precursor *FEMS Microbiol. Lett.* 20:141-144.
- Burlingame, R., and Chapman, P.J. (1983) Catabolism of phenylpropionic acid and its 3-hydroxy derivative by *E. coli* *J. Bacteriol.* 155:113-121.
- Casadaban, M.J., Chou, J., and Cohen, S.N. (1980) *In vitro* gene fusions that join an enzymatically active β -galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals *J. Bacteriol.* 143:971-980.
- Cooper, R.A., and Skinner, M.A. (1980) Catabolism of the 3- and 4-Hydroxyphenylacetate by the 3,4-Dihydroxyphenylacetate pathway in *Escherichia coli* *J. Bacteriol.* 143:302-306.
- Cooper, R.A., Jones, D.C., and Parrot, S. (1985) Isolation and mapping of *Escherichia coli* K12 mutants defective in phenylacetate degradation *J. Gen. Microbiol.* 131:2753- 2757.

- Daumy, G.O., McColl, A.S., and Apostolakis, D. (1982) Repression of Penicillin G of *Proteus rettgeri* by Tricarboxylic Acid Cycle Intermediates *J. Bacteriol.* 152:104-110.
- Denhardt, D.T. (1966) A membrane filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* 23:641-646.
- Deretié, V., Francetic, O., and Gilsin, V. (1984) Instability of the plasmid carrying active penicillin acylase gene from *Escherichia coli*: conditions inducing insertional inactivation. *FEMS Microbiol. Lett.* 24:173-177.
- Gang, D.M., and Shaikh, K. (1976) Regulation of penicillin acylase in *Escherichia coli* by cyclic AMP. *Biochim. Biophys. Acta* 425:110-114.
- García, J.L., and Buesa, J.M. (1986): An improved method to clone penicillin acylase genes: Cloning and expression in *Escherichia coli* of penicillin G acylase from *Kluyvera citrophila*. *J. Biotechnol.* 3:187-195.
- Hamilton-Miller J.M.T. (1966): Penicillinacylase. *Bacteriol. Rev.* 30:761-771.
- Kaufmann, W., and Bauer, K. (1964) The production of penicillin amidase by *Escherichia coli* ATCC9637. *J. Gen. Microbiol.* 35, IV.
- Lehrach, H., Diamond, D., Wozney, J.M., and Boedtker, H. (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* 16:4743-4751.
- Lowry, D., Rosebrough, N., Farr, A., and Randall, R. (1951) Protein measurement with the folin phenol reagent. *J. Biol. Chem.* 193: 265-275.
- Mahajan, P.B. (1984) Penicillin acylases. An update. *Appl. Biochem. Biotechnol.* 9:537-554.
- Margolin, A.L., Svedas, V.K., and Berezin I.V. (1980) Substrate specificity of penicillin amidase from *E. coli*. *Biochim. Biophys. Acta* 616:283-289.

- Mulligan, M.E., Hawley, D.K., Entriken, R., and McClure, W.R. (1984) *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acid Res.* 12:789-800.
- Oh, S.-J., Kim, Y.-C., Park, Y.-W., Min, S.-Y., Kim, I.-S., and Kang, H.-S. (1987) Complete nucleotide sequence of the penicillin G acylase gene and flanking regions, and its expression in *Escherichia coli*. *Gene* 56:87-97.
- Oliver, G., Valle, F., Rosetti, F., Gómez-Pedrozo, M., Santamaría, P., Gosset, G., and Bolivar, F. (1985). A common precursor for the two subunits of the penicillin acylase from *Escherichia coli* ATCC11105. *Gene* 40:9-14.
- Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237-251.
- Savidge T.A. (1984) Enzymatic conversions used in the production of penicillins and cephalosporins. in : Vandame E. J.(ed) *Biotechnology of industrial antibiotics*. Marcel Dekker New York pp. 172-224.
- Schumacher, D., Sizmann, D., Haug, H., Buckel, P., and Böck, A. (1986) Penicillin acylase from *E. coli*: unique relation. *Nucleic Acids Res.* 14:5713-5727.
- Shimi, I.R., and Imam, G.M. (1986) Biological breakdown of Benzyl-penicillin by preformed mats of penicillin producing organisms. *Biochem. J.* 101:831.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Szentirman, A. (1964) Production of penicillin acylase. *Appl. Microbiol.* 12:185.

- Valle, F., Gosset, G., Tenorio, B., Oliver G., and Bolivar, F. (1986) Characterization of the regulatory region of the *Escherichia coli* penicillin acylase structural gene. *Gene* 50:119-122.
- Valle, F., Balbás, P., Merino, E., and Bolivar, F. (1991) The role of the penicillin amidase in Nature and in industry. *Trends in Biological Sciences*. 16 (1991) 36-40.
- Vandame, E.J. (1980) in *Economic Microbiology* Vol. 5 (Rose, A, H., ed.), *Penicillin Acylases and Beta lactamases* pp 467-522, Academic Press.
- Vojtisek, V., and Slezák, J. (1975a) Penicillinamidohydrolase in *Escherichia coli*. II. Synthesis of the enzyme, kinetics and specificity of its induction and the effect of oxygen. *Folia Microbiol.* 20:289-297.
- Vojtisek, V., and Slezák, J. (1975b) Penicillinamidohydrolase in *Escherichia coli*. III. Catabolite Repression, Diauxie, Effect of cAMP and Nature of the Enzyme Induction. *Folia Microbiol.* 20:298-306.
- Young, F.S., and Furano, A.V. (1981) Regulation of the synthesis of *E. coli* elongation factor Tu. *Cell* 24:695-706.

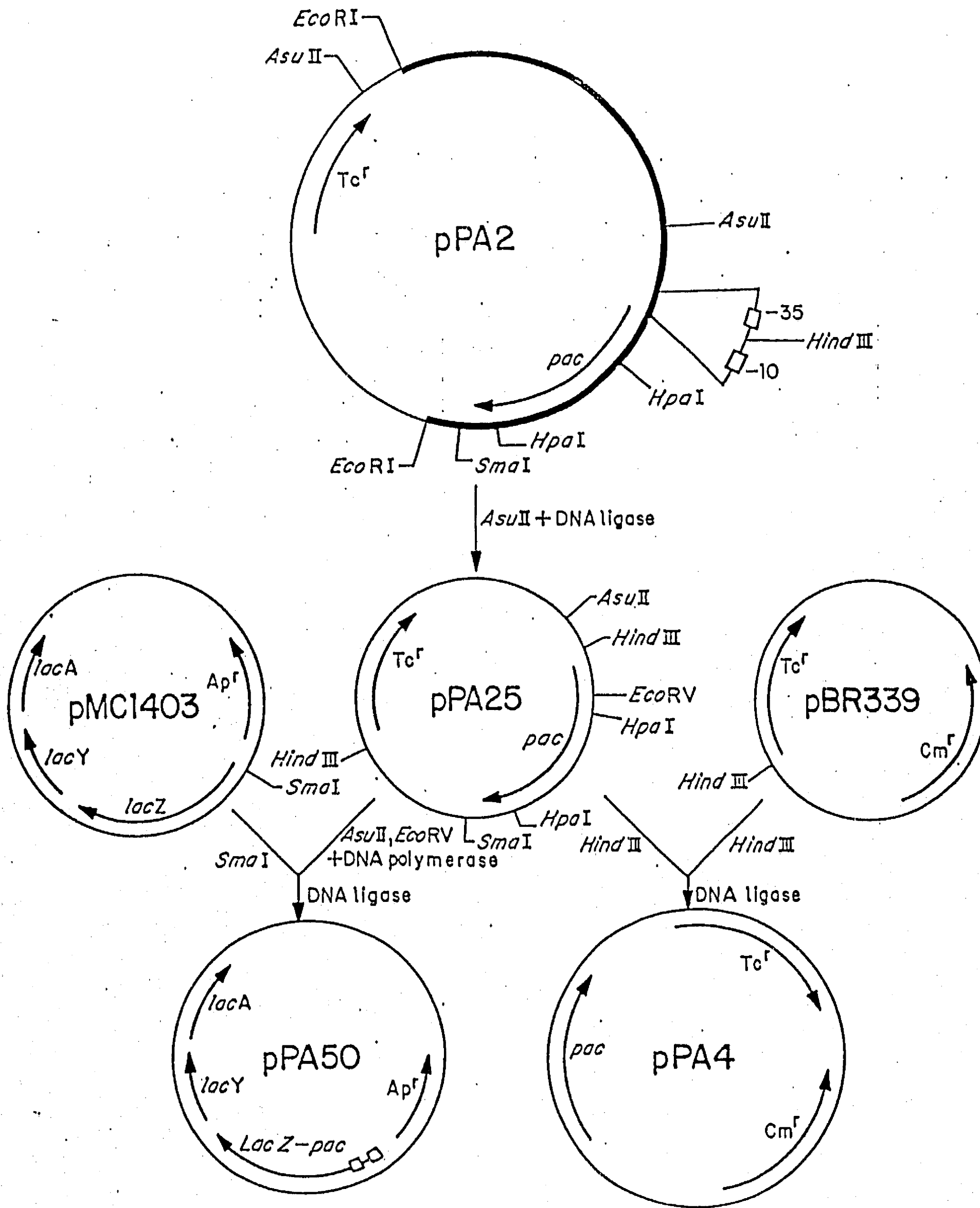
Fig. 1. Strategies for construction of the *pac* derivative plasmids. Plasmids pPA2, pPA4 and pPA25 have been previously described (Oliver *et al.*, 1985). The thick line in plasmid pPA2 represent the *E. coli* ATCC11105 DNA fragment that carries the *pac* gene. The open boxes represents the -35 and -10 regions of the *pac* gene. Plasmid pPA50 is a pPMC1403 (Casadaban *et al.*, 1980) derivative in which the *lacZ* gene has been fusioned to the 5' regulatory region of *pac* gene. Only the relevant restriction sites are indicated.

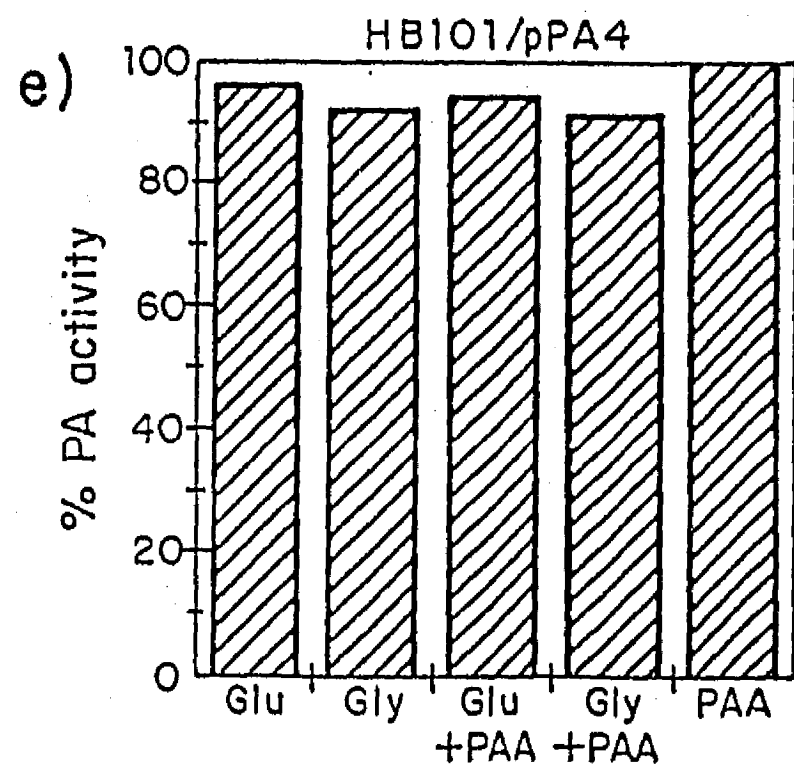
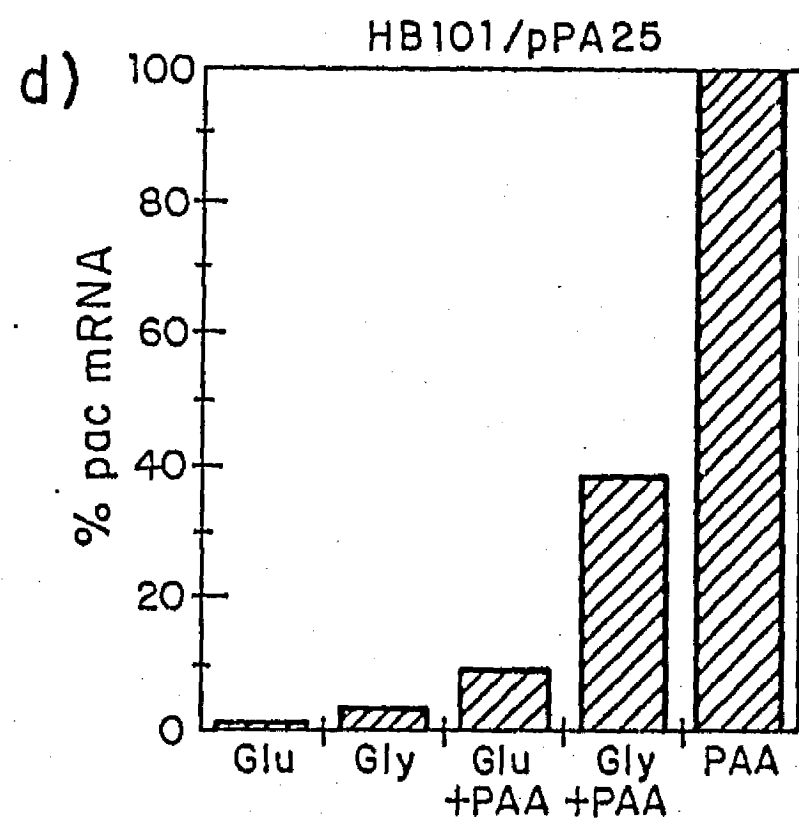
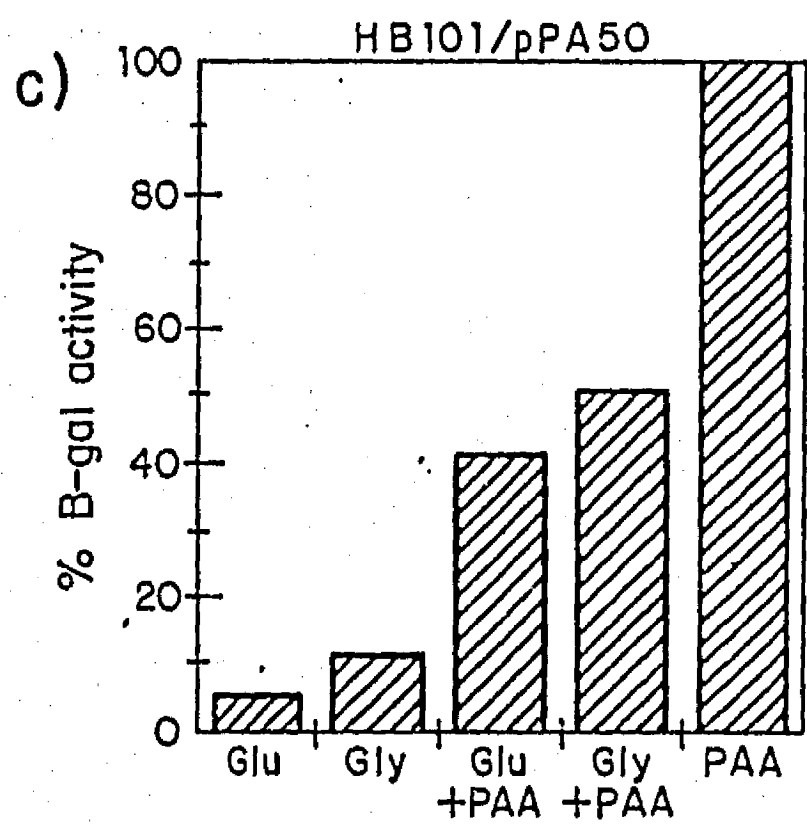
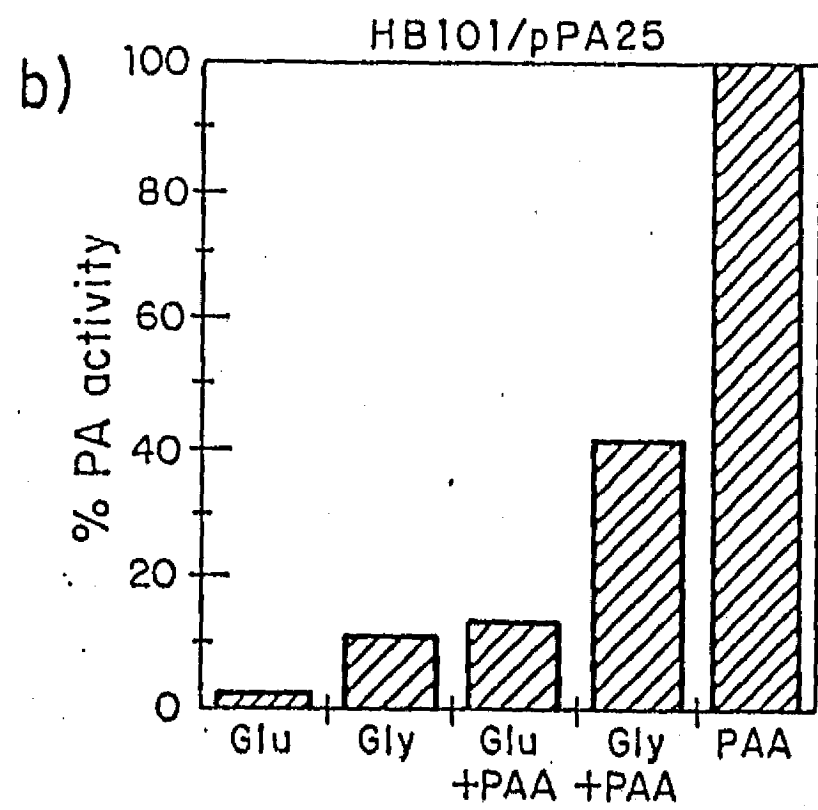
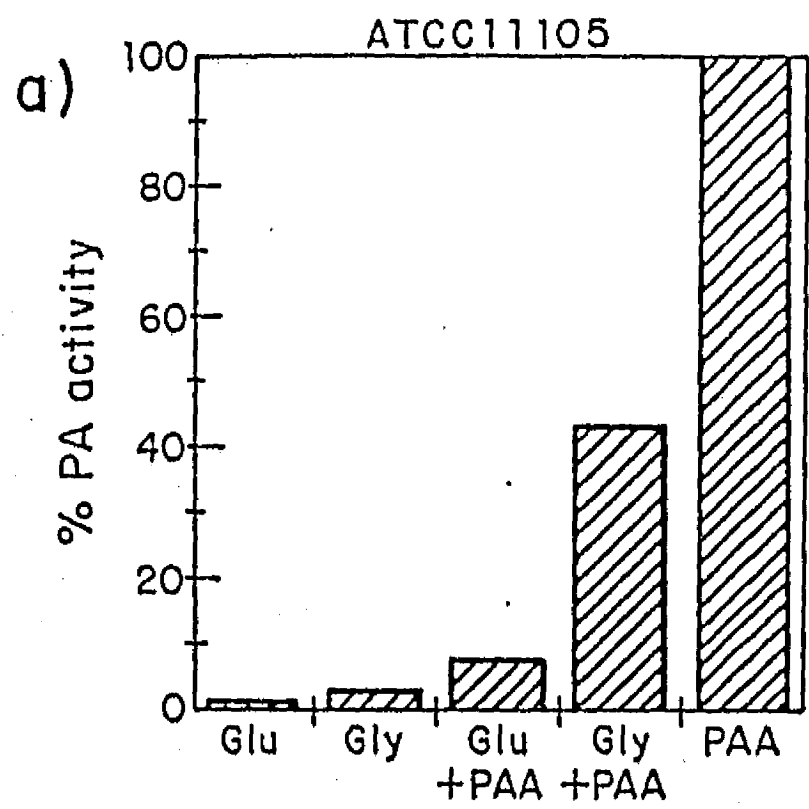
Fig. 2. Profiles of PA, β -galactosidase and *pac* mRNA. *E. coli* ATCC11105 and HB101, transformed with the indicated plasmids, were grown in NN minimal media at 28 °C and harvested at late exponential phase. Five different induction condition were studied: Glu: 11 mM glucose, Gly: 22 mM glycerol, Glu+PAA: 11 mM glucose + 0.1% phenyl acetic acid, Gly+PAA: 22 mM glycerol + 0.1% PAA and PAA: 0.1% PAA. In order to have a comparative analysis of these results, the values are expressed in relative terms respect to the maximum value (100%) of each case.

Fig. 3. Putative promoters in the 5' region of the *pac* gene. a) Histogram of the putative promoters found within the first 160 nucleotides upstream the initiation ATG codon. The statistical algorithm used in this study was the one developed by Mulligan *et al.* (1984). Arrows indicate the cutoff value for where the putative promoters proposed by Oh *et al.* (1987) (A), or Valle *et al.* (1986) (B) are located; the maximum score obtained within this region, corresponds to the latter promoter. b) The 80 bp preceding the initiation ATG codon of the *pac* gene are presented, where the transcriptional startpoint (+1 at open arrow) and the putative ribosome-binding site (RB) are indicated. The -35 and -10 regions of the promoter proposed by Valle *et al.* (1986), are indicated by rectangles.

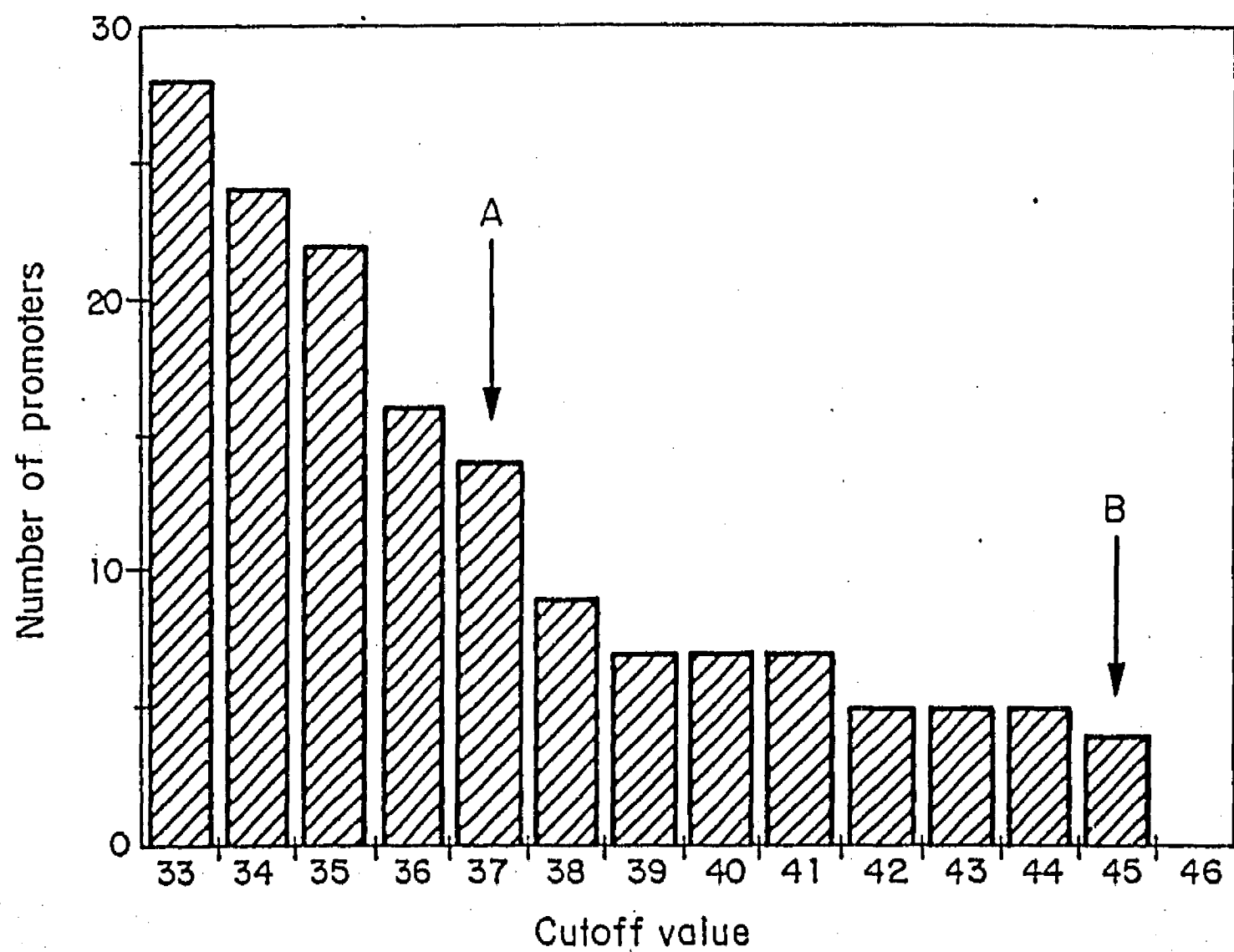
Fig. 4. Southern blot analysis for detection of the chromosomal *pac* gene. Total DNA from *E. coli* strains ATCC11105, HB101, JM101, W3110 (lanes 1, 2, 3 and 4, respectively), were digested with *EcoRI* enzyme, separated by electrophoresis through a 1% agarose gel, transferred to a nitrocellulose membrane and hybridized with a ³²P-labelled *pac* DNA internal fragment (see *Experimental procedures*). Sizes of phage lambda DNA markers are shown in bp.

Fig. 5. Growth of *E. coli* on PenG as carbon source. *E. coli* strain JM101, carrying either plasmid pMMB22 or plasmids pMMB22 and pPA25, was grown at 28 °C on NN minimal media, supplemented with 0.02% casaminoacids and 0.05% of PenG as sole carbon and energy sources. Cell density is expressed as O.D.560

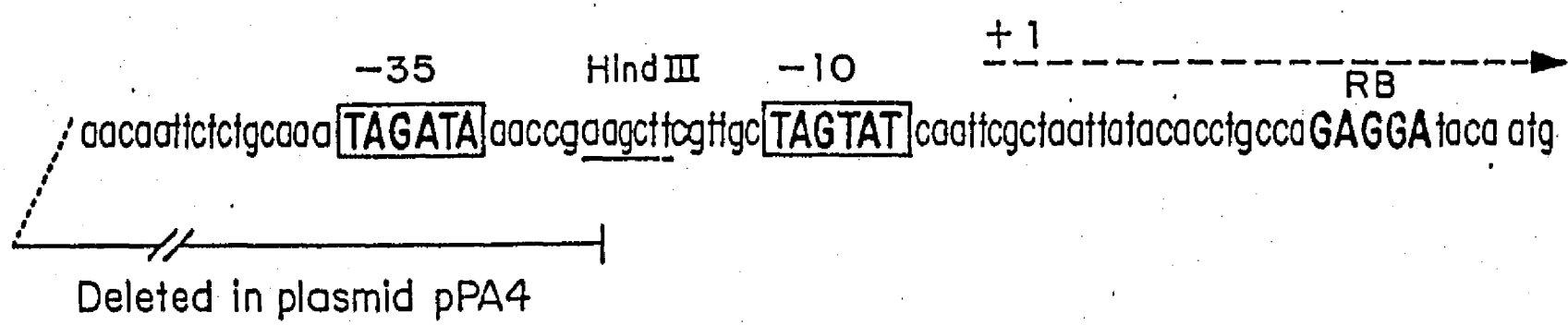




a)



b)



1 2 3 4

23130 -

9416 -

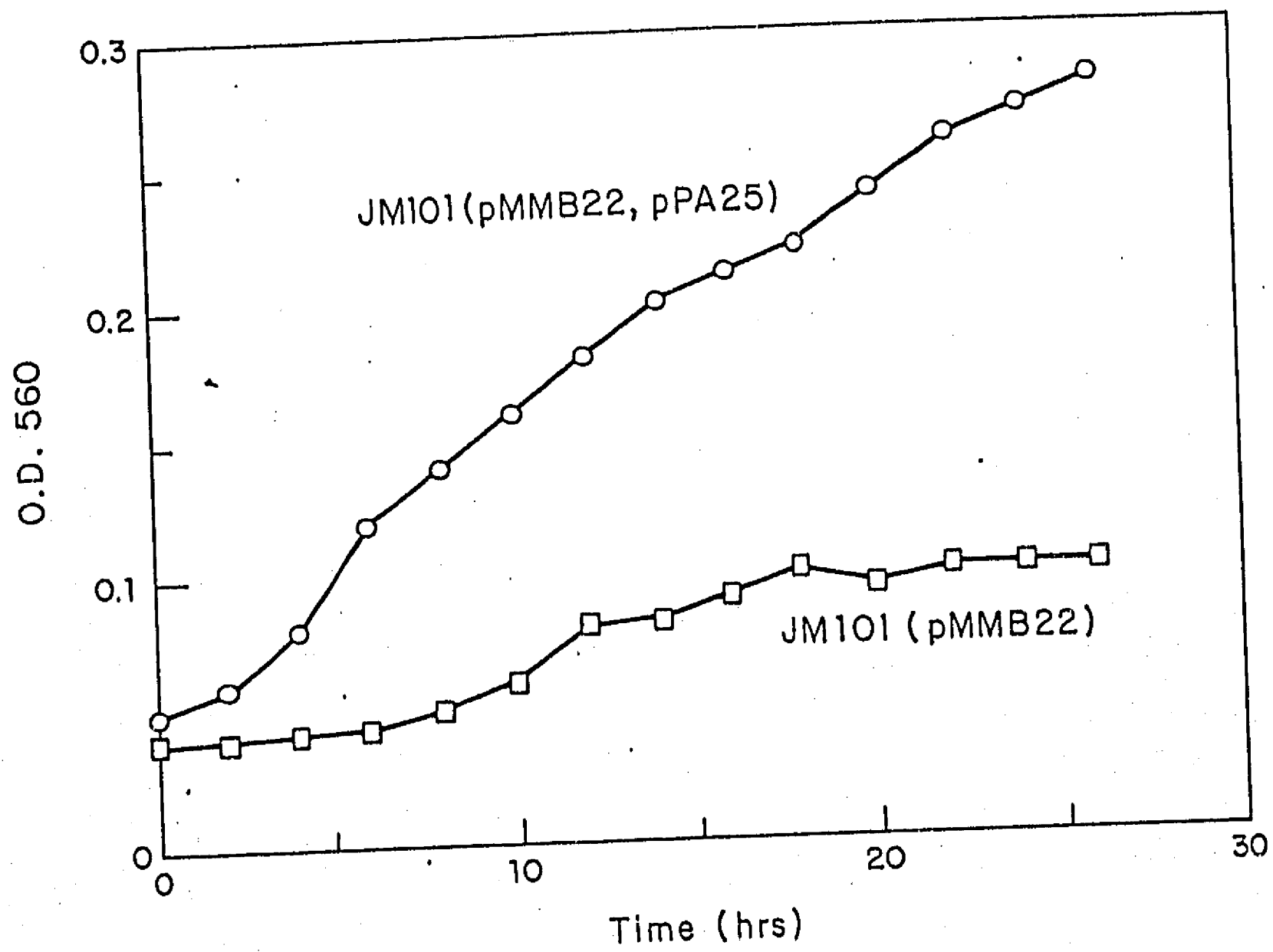
6557 -

4361 -

2322 -

2025 -

File 4



The Journal of Laboratory Technology for Bio-research

Biotechniques

January 9, 1992

Francisco Bolivar, Ph.D.
Universidad Nacional Autonoma de Mexico
Molecular Biology, Instituto de Biotecnologia
Adpo. Postal 510-3
Cuernavaca
Morelos CP 622271
MEXICO

RE: MS # BT91-360

Dear Professor Bolivar:

I am pleased to inform you that your paper entitled "Recovery of DNA from agarose gels..." has been accepted for publication. Based strictly on the date of acceptance, your paper would be scheduled tentatively for the August 1992 issue. Other factors, such as priority recommendations made by reviewers, subject matter appropriate for a particular issue, or the number of papers that can be published in each issue, will determine whether or not your paper can be published earlier. You will be informed as soon as a more definitive publication date has been set.

Enclosed is a form that you are asked to return, within two weeks, if you do not want your paper to be made available through BioTechNet prior to actual publication in *Biotechniques*. BioTechNet is our international computer network for research biologists. Most papers accepted for publication are listed in the network database as articles "in press" and are available for a nominal fee as copies of the accepted manuscript or copyedited galley proofs.

Thank you for this fine submission and for your interest in our journal.

Sincerely,



Thomas J. Stevens
Assistant Editor

RECEIVED
Merrill et al.

PLEASE CHECK GALLEY PROOFS VERY CAREFULLY
AUTHORS ASSUME FULL RESPONSIBILITY FOR CORRECTIONS.
APPROVES AS IS ☐ OR WITH CORRECTIONS INSERTED ☐
AUTHOR'S INITIALS: _____ DATE: _____

A General, PCR-Based Method for Single or Combinatorial Oligonucleotide-Directed Mutagenesis on pUC/M13 Vectors

A method for oligonucleotide-directed, site-specific mutagenesis is reported that represents an extension of methods based on PCR. The scheme affords the following advantages: first, it relies on a "universal" set of primers, complementary to the areas that flank the cloning region of pUC/M13 vectors, thus requiring only one additional oligonucleotide per target site; and second, it is suitable for generating simultaneous, noncontiguous mutations on the same molecule with high efficiency. The method is simple and robust, and uniquely suited for the generation of combinatorial libraries of nonadjacent mutations.

Recent literature shows a number of experimental approaches for oligonucleotide (oligo)-directed, site-specific mutagenesis employing PCR (2-4, 6,7,9). Some important advantages of these approaches are high yield of mutant recovery, with concomitant overall speed of the procedure, and easy applicability on double-stranded templates. Unfortunately, many of the procedures proposed to date require the synthesis of several oligos for each mutagenesis experiment. In this report, we describe a set of oligos that have been successfully utilized to perform mutagenesis experiments, in a manner similar to that described by Nelson and Long (6), but applied to DNA fragments cloned in M13 and pUC vectors. In our opinion, this approach is an attractive alternative to other schemes that also call for subcloning of the fragment to be mutagenized (1,10,11). We consider the method particularly suited for random, targeted mutagenesis because of the capability of the PCR to proceed with oligos complementary to the template only at their 3' ends. Furthermore, as we show in this report, it is also possible to use this method in a serial manner to perform combinatorial

mutagenesis over several, noncontiguous sites. The use of the additional "tail" (oligo A, Figure 1) that is not complementary to the original template sequence, allows combining each PCR mutant product with the mutation encoded by a successive oligo, with avoidance of nonmutant sequences.

The explosive growth in the generation of new methods based on the PCR suggest caution in calling a method "novel" or "unique." Approaches such as those proposed by Sarkar and Sommer (9) or Higuchi et al. (3) share several features with the method described here, and they may be simpler to perform in some cases. Still, to the

best our knowledge (derived from careful monitoring of the directed-mutagenesis literature), the approach described here is general, simple and less costly than previously reported ones. It is also uniquely suited for the introduction of several noncontiguous mutations in the same molecule.

The mutagenesis procedure is described in the Figure 1 legend. The progress of the PCRs was monitored by running samples of them on 7.5% polyacrylamide gel electrophoresis. As shown in Figure 2, products of the expected length were the major species at each stage. The final PCR product was digested with *Pst*I and *Hind*III, ligated

Authors are responsible
for correct references.
They will not be checked
at the publishing office.

Dear Authors,
Please cite ref 5 in text. Thank

Figure 1. Mutagenesis procedure. Plasmid M13mp19PHEco was used for mutagenesis. This plasmid is a derivative of M13mp19, in which a *Pst*-*Hind*III fragment of the gene encoding endonuclease *Eco*RI (10) was cloned. Mutagenesis was designed to introduce two nonadjacent restriction sites in the same molecule. Five oligos were used, A-E. Oligo A (GGAATAGCCGATCGAGATCTAG-GAAACAGCTATGACCATG), Oligo B (GGAATAGCCGATCGAGATCTAGGA) and Oligo C (GTTGTAAACGACGCGCCAGT) are based on sequences adjacent to the M13mp and pUC series of vectors (they are actually extensions of "universal" and "reverse" M13 sequencing primers designed to have similar Tms) and should be able to function with any DNA cloned on them. Oligos D (GCCAT-TAGATCTTGATCTCCTC) and E (TTAACAACGCGTCCATCTGGTC) incorporate *Bgl*II and *Mlu*I sites, respectively (altered bases, with respect to the original gene sequence, are boldfaced). A 100- μ l reaction was set up with 1 ng of M13DNA (digested with *Pvu*II to allow for heteroduplex formation), combined with 30 pmol each of the oligos A and D and subjected to PCR with the following conditions: 30 cycles of 1 min at 93°C, 1 min at 55°C and 1 min at 70°C, followed by a final reaction of 15 min at 70°C (reaction 1). PCR product I was gel-purified and 10 ng of it were used next, in a 90 μ l reaction, together with 10 ng of template (*Pvu*II-digested M13 clone) performing 5 cycles of 1 min at 93°C, 1 min at 50°C and 1 min at 70°C (reaction 2). The resulting extension product was amplified by adding 30 pmol each of oligos B and E to the same tube, bringing the total volume to 100 μ l and performing 27 additional cycles, identical to the first five (reaction 3). Note that oligo B is not complementary to the M13 template and, therefore, should only amplify DNA derived from the primer extension of product I, once the complementary strand is afforded by primer E. The resulting PCR product II was used next, in exactly the same conditions as the previous reaction, with the primer extension product amplified by oligos B and C.

to M13mp19 digested with the same two enzymes and transformed into *E. coli* DH5α™.

Twenty-four independent white plaques were picked and verified for the presence of *Bgl*II and *Mlu*I sites by restriction analysis. Twenty-three clones had both sites incorporated (the other clone was lacking an insert). We also determined the nucleotide sequence of one of the clones and confirmed that it corresponded to what was expected, with the predicted changes at the target sites for mutagenesis and no other alterations.

We have repeatedly used oligos A, B and C to perform other mutagenesis experiments and also oligo B as a companion to similar schemes on different templates (that is, using a corresponding oligo, analogous to A). We have found the approach fast and reliable, and the only viable alternative to perform combinatorial mutagenesis on nonadjacent sites. We have also extended this procedure to general five nonadjacent point mutations, encompassing over 1000 bases, with satisfactory results (data not shown).

Additional mutations, not encoded by the mutagenic oligos, have been observed in our laboratory, at rates consistent with previous reports (3). To keep this drawback at a minimum, we suggest using gene fragments as short as possible and trying new, less error-prone, thermally stable polymerases as they become available.

REFERENCES

1. Carter, P., H. Bedouille and G. Winter. 1985. Improved oligonucleotide site-directed mutagenesis using M13 vectors. *Nucleic Acids Res.* 13:4431-4443.
2. Giebel, L.B. and R.A. Spritz. 1990. Site-directed mutagenesis using a double-stranded DNA fragment as a PCR primer. *Nucleic Acids Res.* 18:4947.
3. Higuchi, R., B. Krummel and R.K. Saiki. 1988. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* 16:7351-7367.
4. Ho, S.N., H.D. Hunt, R.M. Horton, J.K. Pullen and L.R. Pease. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59.
5. Kuhn, I., F.H. Stephenson, H.W. Boyer and P.J. Greene. 1986. Positive selection vectors utilizing lethality of the *Eco*RI endonuclease. *Gene* 44:253-263.
6. Nelson, R.M. and G.L. Long. 1989. A general method of site-specific mutagenesis using a modification of the *Thermus aquaticus* polymerase chain reaction. *Anal. Biochem.* 180: 147-151.
7. Perrin, S. and G. Gilliland. 1990. Site-specific mutagenesis using asymmetric polymerase chain reaction. *Nucleic Acids Res.* 18:7433-7438.
8. Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
9. Sarkar, G. and S. Sommer. 1990. The "megaprimer" method of site-directed mutagenesis. *BioTechniques* 8:404-407.
10. Sayers, J.R., W. Schmidt and F. Eckstein. 1988. 5'-3' exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res.* 16:791-802.
11. Zoller, M. and M. Smith. 1984. Oligonucleotide-directed mutagenesis: A simple method using two oligonucleotide primers and a single-stranded DNA template. *DNA* 3:479-488.

We are grateful to Paul Gaytán for oligonucleotide synthesis and to Sonia Caro for secretarial assistance. This work was supported by a grant from DGSCA, National University of México. Address correspondence to X. Soberón.

Enrique Merino, Joel Osuna,
Francisco Bolívar and Xavier
Soberón
CIIGB, UNAM
Apdo. Postal 410-3
62271 Col. Miraval
Cuernavaca, Morelos, Mexico

For comments or questions, you may contact the author at the BioTechNet E-Mail address below.

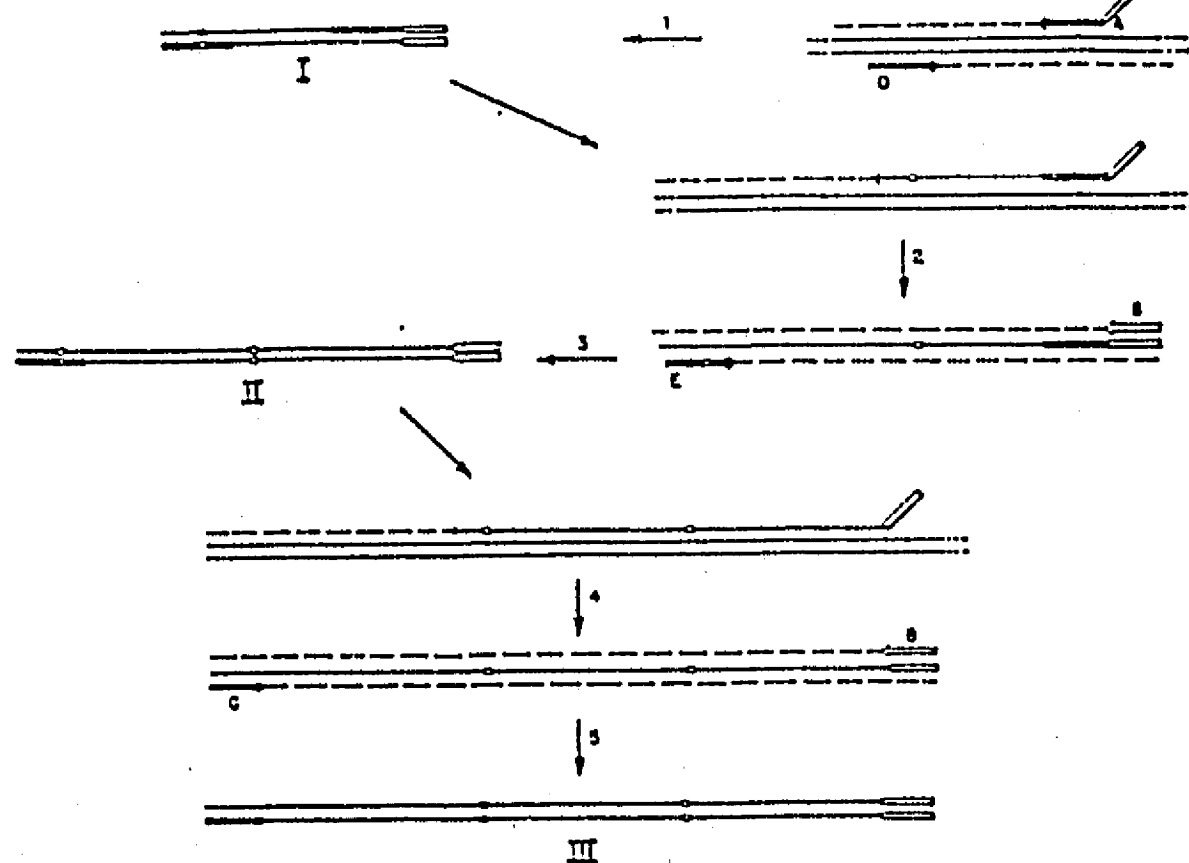


Authors are responsible
for correct references.
They will not be checked
at the publishing office.

Author Please Note: The BioTechNet logo you see below will appear at the end of your article with your BioTechNet E-Mail address should you elect to take advantage of the offer presented elsewhere in the materials received with your galley. Please provide a response.

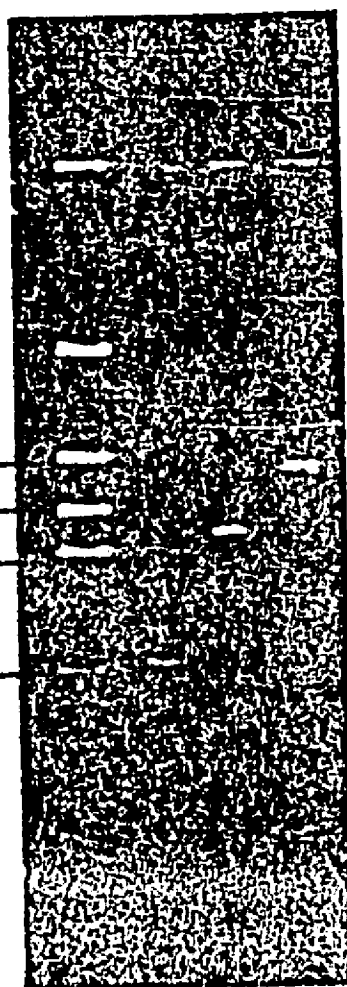
Figure 2 Electrophoretic analysis of the PCR products. A, size markers; B, C and D, PCR products I, II and III (see Figure 1), respectively.

61 91-202
Murino



F2
BT91-202
Murino

587
458
434
267



A B C D

Biotechniques

February 13, 1992

Xavier Soberon, Ph.D.
Universidad Nacional Autonoma de Mexico
UNAM; Adpo. Postal 510-3
Cononia Miraval 62271
Curenavaca, Morelos
MEXICO

RE: MS # BT91-202

Dear Doctor Soberon:

Enclosed is the galley proof of your article entitled "A general, PCR-based method for single or combinatorial...".

Please read the galley proof VERY CAREFULLY. Indicate if you disagree with any editing changes, making any necessary corrections directly on the galley proof. Our copy editor may have made certain changes in wording, sentence structure and terminology where it was felt that readability could be improved without changing the technical meaning.

Please indicate your approval of the galley proof without further correction ("as is") or "with corrections inserted" and return the corrected galley proof to me by February 21, 1992 so that we may include your paper in the upcoming April 1992 issue. We will need the original hard copy of the galley with a cover letter explaining any corrections. Should you need to FAX the galley to us, please send the hardcopy and cover letter by mail. When FAXing, do not use a highlighting pen nor write in any of the margins as these details will not be transmitted. Be sure to program your FAX to its finest setting.

I am also enclosing information concerning reprints, which may be ordered at your convenience. Prepayment is required on all orders to be shipped outside of the U.S. Please call for shipping charges. All payments must be in U.S. funds, drawn on a U.S. bank.

Thank you again for your fine submission.

Sincerely,



Thomas J. Stevens
Assistant Editor

RECOVERY OF DNA FROM AGAROSE GELS
STAINED WITH METHYLENE BLUE.

Address correspondence to E. Merino

Noemi Flores, Fernando Valle, Francisco Bolivar and Enrique Merino.

Department of Molecular Biology, Instituto de Biotecnología,
Universidad Nacional Autónoma de México.

Apdo. Postal 510-3, Cuernavaca Mor. CP 62271, México.

Tel. (52)(73)17-2399. FAX (52)(73)17-2388.

RECOVERY OF DNA FROM AGAROSE GELS STAINED WITH METHYLENE BLUE.

Numerous methods have been published for isolating and purifying DNA from agarose gels. However, most of them utilize ethidium bromide (EtBr) staining and UV light for DNA visualization. This method has some disadvantages due to the carcinogenic effects of EtBr, complicating handling and disposal of this chemical and stained gels. Furthermore, UV radiation also induces mutations that might affect integrity of the isolated DNA. In this communication we report the application of methylene blue (MB) staining to visualize DNA in agarose gels. This DNA can be easily isolated and used for ligation and transformation.

To test the sensitivity of our staining method, 4 μ g of pKK223-3 plasmid were digested with *Eco*RI accordingly to the supplier (Boehringer Mannheim). Serial dilutions of this DNA varying from 1 μ g to 65 ng were loaded in two 1% agarose-TBE gels (5 mm thick with 2x6 mm wells). One gel was stained with EtBr using a standard protocol (2). The other gel was stained for 15 minutes with a 0.02% MB solution in distilled water. Afterwards, the gel was de-stained with distilled water for 15 minutes.

The EtBr-stained gel was photographed using a UV transilluminator and the MB-stained gel was photographed using a white light box. The results are presented in figure 1. As can be seen, using MB-staining, 62.5 ng of DNA can be visualized easily (see fig. 1B). However, EtBr is four times more sensitive than MB, therefore the MB staining procedure is mainly applicable when high amounts of DNA are loaded, i.e. preparative gels. It is important to mention that longer de-staining times does not increase resolution.

In order to compare the transformation efficiency of MB with EtBr-stained DNA, 4 µg of M13 mp8 DNA phage digested with *EcoRI* were loaded into two gels. They were electrophoresed and stained with EtBr or MB respectively. Afterwards, the DNA bands were visualized with the appropriate light source, cut out with a razor blade and placed into microcentrifuge tubes. Two different methods were used to purify the DNA from agarose. The phenol-freeze method (1) and NaI-glass method (accordingly to the suppliers of the GeneClean kit, Bio 101, La Jolla, C.A.). 0.1 µg of DNA isolated by these two methods, were ligated and transformed into calcium-chloride competent *E. coli* JM101 cells. The number of plaques were scored and the results are presented in Table 1. The above procedure was performed three times from which reproducible results were obtained.

The number of plaques obtained with MB stained-DNA was twice of EtBr-stained DNA and did not depend on the purification method. Based on these results, we believe that the MB staining procedure offers a good and safe alternative to obtain a highly clonable DNA, even that, when it is compared with EtBr, is less sensitive.

REFERENCES

1. Benson, S.A. 1984. A rapid procedure for Isolation of DNA Fragments from Agarose Gels. *BioTechniques*. 2:66-67.
2. Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. *Molecular Cloning: A laboratory Manual*, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Table 1.

Staining and isolation procedures	No. of plaques x10 ⁻³
--------------------------------------	-------------------------------------

MB/NaI-glass	3.0
--------------	-----

EtBr/NaI-glass	1.4
----------------	-----

MB/Phenol-freeze	3.2
------------------	-----

EtBr/Phenol-freeze	1.7
--------------------	-----

Table 1. Number of plaques obtained after transforming *E. coli* with DNA isolated from agarose gels stained with MB or EtBr. M13 mp18 DNA was digested with *EcoRI* and electrophoresed into a 1% agarose-TBE gel. Different lanes were stained with either MB or EtBr. After purification using either NaI-glass or phenol-freeze methods, 0.1 µg of these DNAs were ligated and used to transform *E. coli* JM101 cells. The number of plaques obtained with each one of the DNAs are indicated.

Figure 1. DNA stained with MB or EtBr. Serial dilutions of M13 mp8 linear DNA: 1000, 500, 250, 125, 62.5, 31.25 and 16.25 ng (lanes 1-7, respectively), were loaded and electrophoresed into a 1% agarose-TBE gel and stained with either EtBr (A) or MB (B).

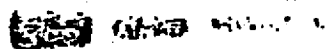
A)

1 2 3 4 5 6 7 8



B)

1 2 3 4 5 6 7 8



ORIGINS OF LIFE AND EVOLUTION OF THE BIOSPHERE

The Journal of the International Society for the Study of the Origin of Life

Editor: JAMES P. FERRIS

Department of Chemistry
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
U.S.A.

Telephone: (518) 276-8493
Telex: 6716050 RPI TROU
BITNET: JFerris@RPITSMTS
Fax: (518) 276-4045

October 11, 1991

Dr. Francisco Bolivar
Dept. of Molecular Biology
Universidad Nacional Autonoma
Apdo. Postal 510-3
Cuernavaca
Mor. CP 62271, MEXICO

Dear Dr. Bolivar:

Re: Ms. #475 - New Insights on the Comma-Less Theory (2nd Revision)

Thank you for providing an additional clarification of your manuscript.
I have accepted it for publication and forwarded it to the publisher. You will
receive page proofs in about two months.

Thank you for your contribution to the Journal.

Sincerely yours,


James P. Ferris
Editor

JPF:sm

NEW INSIGHTS ON THE COMMA-LESS THEORY

Enrique Merino, Paulina Balbás and Francisco Bolivar

**Department of Molecular Biology, Instituto de Biotecnología, Universidad Nacional
Autónoma de México.**

Apdo. Postal 510-3, Cuernavaca, Mor. CP 62271, México. Tel.

(5273) 17 23 99.

Offprint requests to: Dr. Francisco Bolivar

Key words: Genetic code, Comma-less theory, Sequence analysis.

Summary

The comma-less hypothesis represents a theoretical effort to describe one of the steps in the early evolution of the translation apparatus. This hypothesis emphasizes the advantages that a RNY coding pattern would have provided in a primitive RNA adaptor-catalyst system. This theory has been debated for years, both in conceptual and statistical terms, and no consensus about its validity has been ascertained. In this work, a statistical model refuting this theory was reconsidered. This new approach eliminates the bias due to the absence of stop codons in the open reading frame, and to the amino acid composition of bacterial genes. The results obtained support the biological significance of the RNY coding pattern.

Introduction

One of the major questions in molecular biology has been the origin and early evolution of the genetic code. The comma-less theory states that in a primeval stage in the evolution of the translation apparatus, only codons of the form RNY were used, thus providing a primitive message in which translatable codons could only be found in one of the three possible reading frames. This theory emphasizes the positional correlations between purines (R) and pyrimidines (Y), and for years much debate and further elaboration of the concept has been carried out (Crick et al. 1957, 1976; Eigen and Shuster 1978; Shepherd 1981, 1984, 1990; Staden 1984; Wong and Cedergren 1986). With the advent of powerful DNA sequencing techniques and the use of computers to perform searches for reiterative patterns of bases in long stretches of DNA, a prebiotic genetic coding system proposed to be RNY has been supported (Shepherd 1984), although the possibility of the repeating RNY pattern being the outcome of natural selection has not been ruled out (Wong and Cedergren 1986).

Remnants of such RNY-rich messages can still be detected almost everywhere in coding regions. Computer programs have taken into account this property to discern the actual reading frame in newly determined DNA sequences (Shepherd 1981, 1984, 1990). Briefly, in these programs a sequence is scanned so that a test window length of DNA is taken and moved along the sequence in steps. In the test length, Y $\leftrightarrow$ R mutations away from the RNY pattern in the first and third positions are counted in each reading frame, and the RNY codons are considered to be in the frame with the least such mutations.

A hypothesis refuting this notion has been presented by Staden (1984). He claims that although this method does indeed generally indicate the correct reading frame, the biological significance about the RNY code being a relic from a primeval comma-less genetic code is inaccurate. To demonstrate this position, Staden generated very long random amino acid sequences without codon preferences but exhibiting exactly Dayhoff's protein sequence database average amino acid composition. A hypothetical gene was then derived from the amino acid sequence, using the triplets that code for each amino acid in an evenly probable way. He observed that in this gene it was possible to predict the real reading frame with Shepherd's method (1981). From this study Staden concluded that the average amino acid composition and the absence of stop codons in the coding frame, rather than codon preferences, are sufficient to explain how Shepherd's method will choose correctly (Staden 1984). However, this fact alone might not be sufficient to invalidate the comma-less theory because when the information contained in DNA sequences of coding genes is considered, a different outcome is observed.

Materials and Methods

All the programs used in this study were written in C language and ran under a UNIX operating system on a DECstation 3100. The information recorded in the Genbank release 62 was screened prior to the analysis in order to remove redundant gene sequences. The codon values were calculated according to equation 1 and the results are depicted in Table I. The coefficient $V_{\text{RNY}}/V_{\text{non-RNY}}$ was estimated independently for every gene and these results were averaged to obtain the ratio preference.

Results and Discussion

To prove that the information contained in the actual DNA sequences is relevant to the evolutionary theory sustained by the RNY code pattern, we repeated the experiment devised by Staden (1984), using a new approach. Instead of considering the deviations away from the RNY pattern in absolute terms, a numerical value was given to each codon accordingly to the probability of choice between RNY {P(RNY)} versus non-RNY {P(non-RNY)} codons for every amino acid of the randomly generated protein, according to the following equations:

$$(1) \quad V_{\text{RNY}} = 1 - P(\text{RNY}) \quad \text{or} \quad V_{\text{non-RNY}} = 1 - P(\text{non-RNY})$$

Therefore, the value (V) for those codons where no choice is possible between RNY or non-RNY equals zero (ie., the stop codons that are YNR), whereas the value for those codons where a choice is possible, depends on the number of triplets that code for the given amino acid. The value (V) for each codon is presented in Table I.

Following the experimental design by Staden, very long amino acid sequences exhibiting exactly Dayhoff's protein sequence average amino acid composition, were generated at random. From these amino acid sequences, the corresponding nucleotide sequences were also randomly generated and the numerical values for ΣV_{RNY} and $\Sigma V_{\text{non-RNY}}$ were calculated. These values were identical therefore the coefficient $\Sigma V_{\text{RNY}} / \Sigma V_{\text{non-RNY}}$ equaled 1. However, when this analysis was performed on a gene by gene basis, utilizing all the prokaryotic sequences reported in the Genbank, the average coefficient obtained was 1.6. This result implies that even when no bias is introduced due to the presence of stop codons ($V=0$) or to the amino acid composition *per se* (V_{RNY} equals $V_{\text{non-RNY}}$ for every amino acid), a predilection for codons with the RNY form is present.

Most of the statistical analysis concerning RNY pattern on DNA sequences have considered exclusively the RNY frequency of appearance in absolute terms (Staden 1984, Nussimatz 1984, Wong and Cedergren 1986). Our probabilistical approach differs from the previous ones in that the contribution of the RNY pattern is considered relative to its probability of selection and so different outcomes can be seen. The 1.6 coefficient obtained indicated that there is a strong bias in favour of the choice of RNY triplets in present day genes. However, it is not clear whether this fact represents the residue of a primitive gene structure (Shepherd 1984) or the effect of a natural selection process (Wong and Cedergren 1986), or both. Strong argumentation against the view of a primitive genetic coding system has been presented by Wong and Cedergren (1986) based on the analysis of mutational rates of a variety of genes. However, it is important to mention that the theory of the coevolution of the amino acids and the genetic code also developed by Wong (1981), states that the prebiotic amino acids used for primaeval protein synthesis were encoded exclusively by RNY triplets (except Asn). This proposal supports the view that a RNY coding system existed long before life as it is known today. The results obtained in this work taken by their own might not offer sufficient evidence to clear this debated issue. Nevertheless, we consider that our new approach to analyze the data, might help to consider new points of view and is conclusive about the preference to the RNY pattern in present day genes.

Acknowledgments

We thank Dr. Antonio Lazcano and Dr. Xavier Soberón for fruitful discussions and critical review of the manuscript.

References

- Crick, F.H.C., Griffith, J.S. and Orgel, L.E.: 1957, Proc. Natl. Acad. Sci. USA **43**,416-421.
- Crick, F.H.C., Brenner, S., Klug, A., Pieczenik, G.: 1976, Origins Life **7**,389-397.
- Eigen, M. and Shuster, P.: 1978, Naturwissenschaften **65**,341-369.
- Shepherd, J.C.W.: 1981, Proc. Natl. Acad. Sci. USA **78**,1596-1600.
- Shepherd, J.C.W.: 1984, Cold Spring Harbor Symp. Quant. Biol. **47**,1099-1108.
- Shepherd, J.C.W.: 1990, Meth. Enzymol. **183**,180-192.
- Staden, R.: 1984, Nucl. Acids. Res. **12**,551-567.
- Wong, J. T.-F.: Trends Biochem. Sci. **6**,33-35.
- Wong, J.T.-F. and Cedergren, R.: 1986, Eur. J. Biochem. **159**,175-180.

Table I. Numerical values (V) for the genetic code. Triplets of the form RNY are presented in bold face.

	U		C		A		G
U	Phe UUU 0.00	Ser UCU 0.33	Tyr UAU 0.00	Cys UGU 0.00			
	Phe UUC 0.00	Ser UCC 0.33	Tyr UAC 0.00	Cys UGC 0.00			
	Leu UUA 0.00	Ser UCA 0.33	End UAA 0.00	End UGA 0.00			
	Leu UUG 0.00	Ser UCG 0.33	End UAG 0.00	Trp UGG 0.00			
C	Leu CUU 0.00	Pro CCU 0.00	His CAU 0.00	Arg CGU 0.00			
	Leu CUC 0.00	Pro CCC 0.00	His CAC 0.00	Arg CGC 0.00			
	Leu CUA 0.00	Pro CCA 0.00	Gln CAA 0.00	Arg CGA 0.00			
	Leu CUG 0.00	Pro CCG 0.00	Gln CAG 0.00	Arg CGG 0.00			
A	Ile AUU 0.33	Thr ACU 0.50	Asn AAU 0.00	Ser AGU 0.66			
	Ile AUC 0.33	Thr ACC 0.50	Asn AAC 0.00	Ser AGC 0.66			
	Ile AUA 0.66	Thr ACA 0.50	Lys AAA 0.00	Arg AGA 0.00			
	Met AUG 0.00	Thr ACG 0.50	Lys AAG 0.00	Arg AGG 0.00			
G	Val GUU 0.50	Ala GCU 0.50	Asp GAU 0.00	Gly GGU 0.50			
	Val GUC 0.50	Ala GCC 0.50	Asp GAC 0.00	Gly GGC 0.50			
	Val GUA 0.50	Ala GCA 0.50	Glu GAA 0.00	Gly GGA 0.50			
	Val GUG 0.50	Ala GCG 0.50	Glu GAG 0.00	Gly GGG 0.50			

Are overlapping reading frames a remanent of the early coding system?

Enrique Merino, Paulina Balbás, José Luis Puente and Francisco Bolivar.

**Departamento de Biología Molecular, Instituto de Biotecnología,
Universidad Nacional Autónoma de México. Apdo. Postal 510-3,
Cuernavaca, Mor. CP 62271, México. Tel. (5273) 17-23-99.**

Offprint requests to Dr. Francisco Bolivar.

**Key words: Genetic code, open reading frames, comma-less theory,
double stranded coding.**

SUMMARY

We address the question whether the overlapping open reading frames (ORFs) are determined by natural selection or is a relic of the earliest stages in the evolution of the coding system. We performed an extensive search in the Genbank release 63 and identified a large group of genes with totally ORFs in the anti-sense complementary DNA strand. This is a frequent, nonrandom phenomenon which is highly prevalent in bacteria although it is also present in a wide variety of species. The analysis of these bacterial genes with 100% overlapping ORFs contain a higher proportion of triplets in the RNY form (purine-any nucleotide-pyrimidine) as well as higher GC content than the average value obtained for all the genes in the database. The sequences of gene families as well as homologous proteins in different species were studied, but no DNA sequence patterns were observed. The evolutionary implications of these data are discussed in the light of the theoretical capacity of overlapping DNA strands to code for proteins in an early stage of life.

INTRODUCTION

Modern biochemistry provides evidence that metabolism was once based on processes mediated by messenger-catalyst RNA molecules. RNA structures are proposed to have functioned as organizing centers that coordinated the order and nature of biochemical reactions using base-pairing interactions (Tyagi 1981; Gibson and Lamond 1990). There is theoretical and statistical evidence supporting the view that codons of the form RNY (purine - any nucleotide - pyrimidine) were preferentially used in this initial stage, thus providing a primitive message in which translatable codons could mainly be found in one of the three possible reading frames. This could have represented a considerable advantage for a primitive translation system in that a start signal would not be required to indicate the correct reading frame and no blockages of translation would occur due to primaeval RNA adaptor molecules attaching themselves in the wrong frame of the message. This is the basic statement of the Comma-less theory (Crick 1957; Eigen and Shuster 1978; Tyagi 1981; Shepherd 1984, 1990). This theory has been debated both in conceptual and statistical terms, and no consensus about its validity has been ascertained. However, recent

reconsideration of a statistical model strongly refuting the comma-less theory (Staden 1984), has led to new insights that support that the theory might be correct (Merino *et. al.*, 1991, in press).

Most importantly, in a relatively later stage in evolution, when this RNY code was imprinted into double-stranded DNA, these molecules would exhibit the same overlapped, in-phase RNY code pattern in the 5'---> 3' direction, which meant that both strands could have directed the synthesis of different proteins by acting either directly as messengers, or indirectly as templates for RNA synthesis. The discovery of present-day genes with the capacity to code for mRNAs in overlapping strands supports this hypothesis (Rak 1982; Inouye 1988; Simons 88; Henikoff *et. al.*, 1986; Adelman 1987; Spencer *et. al.*, 1986). There are also some reports where large ORFs (Open Reading Frames) have been detected in the antisense strand of a number of genes, but no expression of such putative proteins has been ascertained in most of the cases (Shepherd 1984; Goldstein and Brutlag 1989; Kumamoto and Nault 1989). Moreover, some discussions are available about the theoretical coding capacity of both DNA strands, and the properties that should be exhibited by the proteins coded by both strands (Casino *et. al.*, 1981; Alff-Steinberg 1981; Tramontano 1984; Sharp 1985; Zull and Smith 1990; Miyajima *et. al.*, 1990; Blalock 1990; Brentani 1990; Sloostra and Roubos 1990).

Based on these observations, we decided to conduct a computer analysis of the antisense DNA strands of all the bacterial nucleotide sequences present in the Genbank to evaluate the frequency of occurrence and some properties of these ORFs. Our search detected a prominent group of bacterial genes containing unusually large overlapping ORFs in the antisense DNA strand. A statistical analysis confirmed that the length of these ORFs was far from random (Tramontano 1984) but unexpectedly, a considerable group of genes was identified where the ORF overlapped entirely with the coding gene of the sense strand or even extended further. This phenomenon appears to be prevalent in bacterial organisms for they represent 12% of the total bacterial gene entries in the database, although it is not absent in other species. The genes with totally overlapping ORFs contain a higher preference of triplets in the RNY form, when a choice of RNY versus non-RNY codons is possible (ile, thr, ser, gly, val, ala) than the average value, thus suggesting that the tendency for using RNY codons and the theoretical capacity of coding in both strands may be related. There is also a positive correlation between the GC content of the DNA

sequences analyzed and the overlapping percentage, but this fact alone does not account exclusively for this observation. Some of these genes have been identified as highly expressed and presents an important codon usage bias. We propose a hypothesis which correlates the theoretical two-strand coding capacity to the evolutionary process of the genetic coding system.

MATERIALS AND METHODS

All the programs developed for this study were written in C language and ran under a UNIX operating system on a DECstation 3100. The only source of information used in this study was the one recorded in Genbank release 63. This information was screened prior to the analysis in order to remove redundant nucleotide sequences. Eukaryotic sequences were edited to remove introns. We define an overlapping antisense ORF of a gene as the largest reading frame between two in-phase termination codons TGA, TAG or TAA in the 5'--->3' direction of the antisense DNA strand. When the antisense ORFs extend beyond the initiation or termination triplets of the gene sequence, the length is limited by the coding region of the sense strand. The codon usage for bacterial genes was calculated directly from the information recorded in the database. The GC content of genes was also calculated from the DNA sequences, therefore the values may differ from those obtained experimentally. The RNY index was previously described (Merino *et. al.*, 1991).

RESULTS

Size distribution of antisense ORFs Bacterial sequences of protein encoding genes present in Genbank were used to generate the complementary DNA strand in order to identify the ORFs in every antisense strand in the 5'--->3' direction (antisense ORF). The largest antisense ORF present within each gene was selected to carry out a statistical analysis from which a modal distribution was obtained as it can be seen in Fig. 1a. Most of the antisense ORFs found are distributed in a Gaussian fashion, where the media and the standard deviation are 1059 bp and 701 bp respectively. However, there is a conspicuous group of genes that show a statistically significant antisense ORF length,

which is by different criteria far beyond random.

A control analysis was designed to verify whether the results obtained were significant or an effect of preferential codon usage and amino acid composition. The amino acid sequences of each bacterial protein in the Genbank were used to generate two different hypothetical nucleotide sequence databases, where the corresponding coded proteins are identical to the real ones in order to avoid any bias due to amino acid composition and gene length. The first hypothetical database contained randomly chosen codons for every amino acid (Fig. 1b), whereas the choice of triplets in the second database was generated according to the codon usage exhibited by the actual DNA sequences present in the Genbank (Fig. 1c). The differences between Figs. 1b and 1c are a consequence of codon usage that affect the frequency of the codons TTA, CTA and TCA (stop codons in the antisense strand). However, the difference observed when comparing the hypothetical gene distribution in Fig. 1c with the real gene distribution depicted in Fig. 1a, indicate that the length of these antisense ORFs is not an exclusive consequence neither of codon usage, nor of amino acid composition.

Overlapping percentage of the antisense ORFs The overlapping percentage of every coding gene in the sense strand with its largest antisense ORF was also analyzed, and the results are shown in Fig. 2a. The observed trend matches the one previously observed in Fig. 1a, in that a normal distribution is found for the majority of the genes, while a considerable group of genes have a totally overlapping antisense ORF. In many cases, the antisense ORF extends beyond the initiation or termination codon of the sense strand, and this fact accounts for the size of the peak observed. It is important to note that the lengths of the genes with totally overlapping ORFs are not considerably biased towards smaller sizes (see the insert in Fig. 2a), which supports the view that this is not exclusively a size-dependent phenomenon.

An analogous control analysis was also carried out to verify that this distribution was not fortuitous. Fig. 2b shows the distribution obtained for a hypothetical nucleotide sequence database generated on a gene by gene basis with randomly chosen triplets, as it can be seen, a smaller peak corresponding to randomly generated 100 % overlapping antisense ORFs is observed. The coinciding peak present in Fig. 2c indicates that codon usage imprints a considerable bias in the appearance of overlapping ORFs. Nevertheless, comparison of the peaks in the three graphs in Fig. 2 supports the view that the totally

overlapping antisense ORFs are definitely nonrandom, and are not an exclusive effect of codon usage, amino acid composition or gene size, so therefore they might possess biological significance.

Analysis of the nucleotide sequences of the genes with totally overlapping antisense ORFs. The positional correlation of purines and pyrimidines and the GC content were two parameters evaluated in order to assess the significance of the 100% overlapping antisense ORFs phenomenon. Computer programs were devised to scan for the positional correlations between purines and pyrimidines, and to determine whether or not triplets with the pattern RNY were preferentially used in the genes with 100 % overlapping ORFs. The algorithm has been described in detail elsewhere (Merino *et. al.*, 1991). Suffice it to say that a coefficient indicating the preferential usage of RNY codons versus non-RNY codons whenever a choice is possible (ile, thr, ser, gly, val, ala) was obtained. The bias due to the absence of stop codons in the reading frame and to the average amino acid composition of proteins were eliminated. Absence of codon preference results in a value of 1, so deviations from this figure indicate bias by codon preference. When this analysis was performed on a gene by gene basis in the database, the average coefficient obtained was 1.6, indicating the existence of a predilection for codons with the RNY form. However, when the genes with totally overlapping ORFs were tested in identical conditions, a higher coefficient of 2.2 was obtained. Some of these genes have been identified as highly expressed and are known also to be present in the earliest living organisms.

Higher GC content might result in lower abundance of stop codons in the antisense strand, which could be partially responsible for the prominent peak in Fig. 2a. The overall GC content of the bacterial genes was determined to be 51%. Nevertheless, the genes with totally overlapping ORFs in bacteria contain an average value of 62% GC.

Correlation study of the RNY and GC contents to overlapping percentage of different species. A detailed study of individual organisms was designed in order to determine the correlation between RNY codons and overlapping percentage of individual species. A positive correlation with a value of 0.57 was obtained and the results are presented in Fig. 4.

Determination of the correlation between GC content and overlapping percentage of all the genes reported for those species was also obtained. The average values for both parameters were calculated and plotted in Fig. 3, where a positive correlation of 0.87 can be observed.

DISCUSSION

It is a well established phenomenon that a nonrandom coding capacity exists in the complementary DNA strand of some protein encoding genes (Casino 1981; Tramontano 1984; Alff-Steinberger 1984; Sharp 1985). This capacity is given by large overlapping ORFs, and their frequency in bacterial genes is depicted in Fig. 1a. The data presented thereafter extend these observations by calculating the overlapping percentage of these ORFs, by analyzing the coding pattern exhibited by these genes, and by correlating this capacity to the evolutionary implications of this phenomenon.

It has been argued that long antisense ORFs are a consequence of the low usage of codons TTA, CTA and TCA, which are the triplets opposite to the stop codons in the 5'--->3' direction (Staden 1984). This tendency is observed when comparing Figs. 2b (no codon bias present) and 2c (bias inflicted by bacterial codon usage). However, the fact that such a large number of genes contain a 100 % overlapping single ORF (Fig. 2a), suggests that the function of the antisense DNA strand in a previous evolutionary stage was not solely restricted to serve as a necessary conduit for double stranded semiconservative DNA replication. In this context, we would like to propose that the double strand coding capacity of DNA molecules may have been advantageous in the initial stage to assay a higher number of potential coding sequences in order to obtain more proteins capable of biological function. Once function was obtained, the DNA coding sequence became "frozen" and entered a process of natural selection in order to become improved, while the antisense ORF remained as a relic of this stage.

Examination of the coding pattern of the bacterial genes in the database showed two important facts. The first is that no marked global preference of GC over AT (51% GC) is present in the bacterial database. Nevertheless, when the analysis was done with the genes containing totally overlapping ORFs, a value of 62% GC was obtained. This bias was also observed when genes from individual species were studied. A positive correlation of 0.87 between GC content and overlapping percentage was observed (Fig. 3).

The second fact observed is the existence of a preferential usage of the RNY codons for those amino acids where a choice between RNY and non-RNY triplets is possible (ile, thr, ser, gly, val, ala). There is a considerable bias towards the use of RNY codons in the bacterial database, described by the 1.6 coefficient value observed. However, the genes containing totally overlapping ORFs have an enlarged coefficient of 2.2, indicating that

these codons are largely preferred. The bias due to preferential usage of RNY codons is also observed when genes from individual species were studied. In this case, a positive correlation of 0.57 between RNY and overlapping percentage was obtained (Fig. 4). The importance of both findings is supported by the observation that measurements of the stability and lifetime of GC and AT pairs imply that in the early phase of primitive translation, GC-rich sequences had the advantage over AT-rich sequences (Kuppers 1984). This is in agreement with the idea that the genes with totally overlapping ORFs retain some properties of the earliest genes, therefore their RNY preference and GC content is expected to be higher.

As it was previously mentioned, the RNY pattern on the sense strand, generates also a RNY pattern on the antisense strand in the 5'--->3' direction. The eminent predilection of codons in the form of RNY might provide a device for selection pressure because of the symmetry granted to both DNA strands. The overlapped, in-phase RNY triplets in both strands might then be a reflection of a primitive structure, not only of the genetic code but of the translation process as well because both strands could act as templates. This does not imply that both strands should be today simultaneously transcribed, but rather supports our hypothesis that DNAs with the capacity to code for the highest possible number of proteins may have had a selective advantage in an earlier stage of molecular evolution.

This conjecture is supported by two independent lines of evidence. The first is the theory of the coevolution of the genetic code and amino acid biosynthesis developed by Wong (1975, 1981). According to this author, not all the 20 amino acids were originally present in the primaeval genetic code, but were accommodated in a series of stages until the actual genetic code was established. Remarkably, the earliest amino acids were those which nowadays retain prevalently the form RNY, and none of them produces a stop codon in the antisense strand. Wong questions the value of the comma-less theory as opposed to natural selection. Based on calculated mutation rates of present-day genes, this author discards the possibility of RNY as relics of an ancient pattern of genetic coding (Wong 1986). However, in the light of the "frozen stages" hypothesis in the evolution of the genetic code presented above, the results of this study indicate that both, a primitive gene structure and natural selection are liable for the widely spread phenomenon of overlapping antisense ORFs. In accordance with this notion is the fact that the RNY message is particularly well preserved in genes which have been required to be heavily expressed

throughout their evolutionary history. The ribosomal proteins and the translation-related protein factors studied in detail reveal that in fact, the RNY pattern in these proteins is highly preserved. Although it is prevalent in bacterial species, it may still be identified in phylogenetically distant species as well.

The second supporting line of evidence is offered by the observations that interacting sites in eukaryotic hormone-receptor proteins are composed of complementary amino acid sequences that might have originated from a single DNA molecule (Goldstein and Brutlag 1989; Zull and Smith 1990; Blalock 1990; Sloostra and Roubos 1990, Brentani 1990). This molecular recognition theory claims that overlapping ORFs for certain eukaryotic hormones have the capacity to code for putative proteins with complementary hydropathy that might direct the hormone-receptor interaction. Further development and experimental verification of this theory are necessary before its validity can be ascertained, but if this mechanism is indeed present in such an evolved protein interaction system, it is suggestive that it might also be present in other species.

In conclusion, the nature of the molecular features now described and the fact that these are so prevalent in bacteria, strongly suggest that this gene organization offered an evolutionary advantage, which may relate to genetic translation and to the coding capacity of double stranded molecules. The information provided here also supports the view that the RNY coding pattern was once used in earlier life systems, and prevails in present day genes, although it is not certain whether they are a relic from a previous stage in the organization and evolution of the genetic code and the translation process or a result of natural selection. Owing to the lack of evidence of these primitive stages, every hypothesis must remain under constant survey so that the extense gap in our empirical knowledge is closed by theory.

Figure 1. Length distribution of the bacterial overlapping ORFs. (a) Actual bacterial gene sequences. (b) Control analysis using a putative database generated with randomly chosen codons for every amino acid. (c) Control analysis with a putative database generated using the codon usage of the actual bacterial Genbank sequences.

Figure 2. Overlapping percentage distribution of the bacterial overlapping ORFs. (a) Actual bacterial gene sequences. The insert shows the size distribution of genes with totally overlapping ORFs. (b) Control analysis using the putative database containing randomly chosen codons for every amino acid. (c) Control analysis using the putative database generated with the codon usage of the actual bacterial Genbank sequences.

Figure 3. Correlation between the overlapping percentage of the ORFs and the GC content of individual species. The GC content was theoretically calculated using the DNA sequences reported. **Fig. 4.** Correlation between the overlapping percentage of the ORFs and the RNY index of individual species.

REFERENCES

Adelman, J.P., Bond, C.T., Douglass, J. and Herbert, E. (1987) Two mammalian genes transcribed from opposite strands of the same DNA locus. *Science* 235:1514-1517.

Alf-Steinberger, C. (1984) Evidence for a coding pattern on the non-coding strand of the *Escherichia coli* genome. *Nucl. Acids Res.* 12:2235-2241.

Blalock, J.E. (1990) Complementarity of peptides specified by 'sense' and 'antisense' strands of DNA. *Trends Biotechnol.* 8:140-144.

Brentani, R. (1990) Information capacity of both DNA strands. *Trends Biochem Sci* 15:463

Casino, A., Cipollaro, M., Guerrini, A.M., Mastrocinque, G. Spena, A. and Scarlato, V. (1981) Coding capacity of complementary DNA strands. *Nucl. Acids Res.* 9:1499-1518.

Crick, F.H.C., Griffith, J.S. and Orgel, L.E. (1957) Codes without commas. *Proc. Natl. Acad. Sci. USA* 43:416-421.

Crick, F.H.C., Brenner, S., Klug, A. and Pieczenik, G. (1976) A speculation on the origin of protein synthesis. *Origins Life* 7:389-397.

Eigen, M. and Shuster, P. (1978) the hypercycle. A principle of natural self-organization. *Naturwissenschaften* 65:341-369.

Gibson, T.J. and Lamond, A.I. (1990) Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* 30:7-15.

Goldstein, A. and Brutlag, D.L. (1989) Is there a relationship between DNA sequences encoding peptide ligands and their receptors? *Proc. Natl. Acad. Sci. USA* 86:42-45

Henikoff, S., Keene, M.A., Fachtel, K. and Fristrom, J.W. (1986) Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. *Cell* 44:33-42

Kumamoto, C.A. and Nault, A.K. (1989) Characterization of the *Escherichia coli* protein-export gene *secB*. *Gene* 75:167-175.

Kuppers, B.O. (1983) *Molecular Theory of Evolution*. Springer-Verlag, Berlin.

Miyajima, N., Horiuchi, R., Shibuya, Y., Fukushige, S., Matsubara, K., Toyoshima, K. and Yamamoto, T. (1990) Two *erbA* homologous encoding proteins with different T3 binding capacities are transcribed from opposite DNA strands of the same genetic locus. *Cell* 57:31-39.

Merino, E., Balbás, P., and Bolivar, F. (1991) New insights on the Comma-less theory.

Otogens of life. In press.

Rak, B., Lusky, M. and Hable, M. (1982) Expression of two proteins from overlapping and oppositely oriented genes on transposable DNA insertion element IS5. *Nature* 297:124-128.

Sharp, P.M. (1985) Does the non-coding strand code? *Nucl. Acids Res.* 13:1389-1397.

Shepherd, J.C.W. (1984) From primaevial message to present day gene. *Symp. Quant. Biol. CSH* 52:1099-1108.

Shepherd, J.C.W. (1984) Fossil Remnants of a primaevial genetic code in all forms of life? *Trends Biochem. Sci.* 7:8-10.

Simons, R.W. (1988) Naturally occurring antisense RNA control- a brief review. *Gene* 72:35-44.

Sloostra, J. and Roubos, E., (1990) Sense-antisense complementarity of hormone-receptor interaction sites. *Trends Biotechnol.* 8:279-281.

Spencer, C.A., Geitz R.D. and Hodgetts, R.B. (1986) Overlapping transcription units in the dopa decarboxylase region of *Drosophila*. *Nature* 322:279-281.

Staden, R. (1984) Measurement of the effects that coding for a protein has on a DNA sequence and their use for finding genes. *Nucl. Acids Res.* 12:551-567.

Tramontano A., Scarlato, V., Barni, N. Cipollaro, M., Franze, A., Macchiato, M.F. and Cascino, A. (1984) Statistical evaluation of the coding capacity of complementary DNA strands. *Nucl. Acids Res.* 12:5049-5090.

Tyagi, S. (1981) Origin of translation: the hypothesis of permanently attached adaptors. *Origins Life* 11:343-351.

Wong, T.F. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA* 72:1909-1912.

Wong, T.F. (1981) Coevolution of genetic code and amino acid biosynthesis. *Trends Biochem. Sci.* 6:33-36

Wong, T.F. and Cedergren, R. (1986) Natural selection versus primitive gene structure as determinant of codon usage. *Eur. J. Biochem.* 159:175-180.

Zull, J.E. and Smith, S.K. (1990) Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem. Sci.* 15:257-261.

A)

B)

C)

