

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Farmacia

Departamento de Microbiología II

**CARACTERIZACIÓN MOLECULAR DE GENES DE
Clostridium acetobutylicum y *Streptococcus pneumoniae*
QUE CODIFICAN PARA PROTEÍNAS CON DOMINIOS DE UNIÓN A COLINA**

Memoria que para optar al Grado de Doctor en Farmacia presenta:

ANA ISABEL RODRÍGUEZ SÁNCHEZ-BEATO

Director de la Tesis

**Dr. José Luis García López
Investigador Científico
Consejo Superior de Investigaciones Científicas**

Madrid, 1995

Quiero expresar mi más sincero agradecimiento a todas aquellas personas que me han ayudado, tanto a nivel científico como personal, en la realización de esta Tesis:

Deseo comenzar por mi director de Tesis, el Dr. Jose Luis Garcia, no sólo por la formación científica que de él he recibido sino por su disponibilidad para con todo el mundo, no sólo en el ámbito profesional sino en el humano. Es por ello que me siento enormemente afortunada de haber trabajado bajo la dirección de una persona como él, por la que siempre sentiré una profunda admiración y respeto.

No quiero olvidar a la segunda persona que iba a dirigir esta Tesis, la Dra. Concepción Ronka, que aunque actualmente no trabaja con nosotros, siempre me brindó su ayuda, cariño y amistad en los inolvidables comienzos.

Quiero mostrar también mi más profundo agradecimiento al Profesor Rubens López, no sólo por haberme aceptado en su laboratorio sino por sus incondicionales consejos y por el estímulo y cariño que siempre me ha brindado y que me han permitido tener más confianza en mi trabajo. Al Dr. Pedro Garcia por ofrecermé en todo momento, además de su amistad y su apoyo científico y moral, sus valiosas enseñanzas y sus críticas cuando fue necesario. Destacar especialmente también al Profesor Ernesto Garcia por su disponibilidad desinteresada en todo momento y por su admirable calidad humana y profesional que no deja de sorprenderme en el día a día.

También deseo enormemente expresar mi agradecimiento a todos mis compañeros de laboratorio, tanto los que continúan aquí como los que ya marcharon, cuyo nombre sobra citar pues sólo ellos saben a quienes me refiero y, con los que he compartido no sólo el trabajo sino, la amistad, el ocio, las risas, las penas, las broncas (de los jefes) y un sin fin de momentos que sería imposible enumerar.

Nombrar también la colaboración de Esperanza Troyano y la excelente asistencia técnica de Elvira Cano, Manuel Carrasco, Paquita Morante, Aurelio Hurtado, Ricardo Uña, que han facilitado la realización de este trabajo.

Al Profesor César Nombela por su gran cordialidad y su labor de tutoría en la Facultad de Farmacia.

Sin olvidar, por supuesto, a todos mis amigos y, en especial a los miembros de la N.B.A y M.B.A, por el estímulo y confianza que siempre me han demostrado.

Por último quiero agradecerles a mis padres, a quienes debo toda mi formación educativa y, a Miguel por haberme estimulado y aguantado en estos últimos y difíciles momentos.

ABREVIATURAS Y SÍMBOLOS.

A	:	deleción
Ap	:	ampicilina
ATP	:	adenosina 5' - trifosfato.
BrEt	:	bromuro de Etidio
Cl	:	curio
C'm	:	cloranfenicol
CPM	:	cuentas por minuto.
D.O.	:	Densidad óptica
dATP	:	desoxiadenosin 5' -trifosfato
dCTP	:	desoxicitidina 5' - trifosfato.
DEAE	:	diethylaminoetanol.
DNA	:	ácido desoxirribonucleico.
dNTP	:	desoxirribonucleósido 5'- trifosfato.
DOC	:	desoxicolato sódico.
EA	:	etanolamina
EDTA	:	etiléndiaminotetracetato
Em	:	eritromicina
g	:	aceleración de la gravedad.
HP	:	perfil de hidrofobicidad
IPTG	:	isopropil-β-D-tiogalactopiranosido.
IS	:	secuencia de inserción
kb	:	1.000 pb.
kDa	:	kilodaltons
Klenow	:	fragmento klenow de la T4-DNA-polimerasa
L.B	:	medio de cultivo de Luria y Bertani.
Ligasa	:	DNA ligasa del fago T4.
LMF	:	agarosa de bajo punto de fusión.
LRRs	:	repeticiones ricas en leucina
LTA	:	ácido lipoteicoico.
MLS	:	macrólidos, lincomicina, estreptogramina
M_r	:	peso molecular relativo.
mRNA	:	ácido ribonucleico mensajero
N	:	medida de turbidimetría (Nefelos)
ORF	:	marco de lectura abierto.
p/v	:	relación peso/volumen
PAGE	:	electroforesis en geles de poliacrilamida.
pb	:	pares de bases.
PCR	:	reacción en cadena de la polimerasa
PFGE	:	electroforesis en campo pulsado
PI	:	punto isoelectrico
R	:	resistencia.
RBS	:	sitio de unión al ribosoma.
RI	:	inhibidor de ribonucleasa
RNA	:	ácido ribonucleico.
RNA_{sn}	:	ribonucleasa.

SDS	:	dodecil sulfato sódico.
Tc	:	tetraciclina
Tris	:	2-amino-2-hidroximetilpropano-1,3,-diol.
tRNA	:	ácido ribonucleico de transferencia
UFC	:	unidades formadoras de colonias.
UV	:	ultravioleta
v/v	:	relacion volumen/volumen.

INDICE

I- INTRODUCCION.	1
1- MICROORGANISMOS QUE CONTIENEN COLINA EN LA ENVOLTURA CELULAR.	2
1.1- <i>Streptococcus pneumoniae</i>	2
1.1.1- Importancia desde el punto de vista clínico y biológico	2
1.1.2- Aspectos estructurales de las envolturas de neumococo.	3
1.1.2.1- 1.a pared celular.	3
1.1.2.2- 1.a membrana plasmática	6
1.1.2.3- 1.a cápsula	8
1.2- <i>Clostridium acetobutylicum</i>	8
1.2.1- Propiedades de <i>C. acetobutylicum</i> y clasificación	8
1.2.2- Importancia industrial: la fermentación acetobutílica (AB)	11
1.2.3- Aspectos estructurales de la pared de <i>C. acetobutylicum</i>	14
1.3- Otros microorganismos	14
2- EL DOMINIO DE UNIÓN A COLINA: PROTEÍNAS CON CAPACIDAD DE UNIÓN A COLINA	16
2.1- Autolisinas de neumococo	16
2.1.1- La amidasa LytA	17
2.1.1.1- Proceso de conversión	18
2.1.1.2- Inhibidores de la actividad catalítica	19
2.1.1.3- Regulación celular.	20
2.1.2- Características de la glucosaminidasa	21
2.2- Características de la PspA	22
2.3- Enzimas líticas de fagos que infectan <i>S. pneumoniae</i>	24
2.3.1- Estructuración en dominios de las enzimas líticas de <i>S. pneumoniae</i> y sus bacteriófagos	26
2.4- Enzimas líticas de pared de <i>C. acetobutylicum</i>	27
3- TEORÍA MODULAR DE LA EVOLUCIÓN DE LAS PROTEÍNAS	30
4- OTRAS PROTEÍNAS CON MOTIVOS SIMILARES A LOS DE UNIÓN A COLINA	34
4.1- Estructura primaria de las toxinas de <i>C. difficile</i>	34
4.2- Estructura primaria de las glicosiltransferasas y proteínas de unión a glucanos de <i>Streptococcus</i>	37
5- OBJETIVOS.	40
II- MATERIALES Y METODOS	41
1- ESTIRPES BACTERIANAS, BACTERIOFAGOS Y PLÁSMIDOS	42
2- MEDIOS Y CONDICIONES DE CULTIVO	44

2.1- <i>E. coli</i>	45
2.2- <i>S. pneumoniae</i>	45
2.3- <i>C. acetobutylicum</i>	45
3- PROCEDIMIENTOS DE TRANSFORMACIÓN GENÉTICA	46
4- ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA	46
5- OBTENCIÓN DE EXTRACTOS CRUDOS DE CULTIVOS CELULARES	47
6- PURIFICACIÓN DE LA PROTEÍNA CspA.	47
6.1- <i>C. acetobutylicum</i>	47
6.2- <i>E. coli</i>	48
7- ENSAYOS DE ACTIVIDAD ENZIMÁTICA	48
7.1- Ensayo enzimático de los extractos obtenidos de <i>E. coli</i>	48
7.2- Ensayo enzimático de los extractos obtenidos de <i>S. pneumoniae</i>	49
7.3- Ensayo enzimático de las proteínas purificadas	49
8- TÉCNICA DEL WESTERN BLOT	49
9- OBTENCIÓN DE ANTICUERPOS ANTI-CspA	50
10- FRACCIONAMIENTO SUBCELULAR EN <i>C. acetobutylicum</i> NCIB 8052	50
11- ANALISIS DE CARBOHIDRATOS	51
12- DETERMINACIÓN DE LA SECUENCIA N-TERMINAL DE LA PROTEÍNA CspA	51
13- EXTRACCIÓN DE DNA CROMOSÓMICO	51
13.1- <i>C. acetobutylicum</i>	51
13.2- <i>S. pneumoniae</i>	52
14- AISLAMIENTO DE RNA	
14.1- Aislamiento de RNA de <i>C. acetobutylicum</i> NCIB 8052	52
14.2- Aislamiento de RNA de <i>S. pneumoniae</i>	53
15- ELECTROFORESIS DE DNA	53
15.1- Electroforesis convencional.	53
15.2- Electroforesis en campo pulsado (PFGE)	54
16- PREPARACION DE PLASMIDOS	54
17- MANIPULACIÓN DEL DNA CON ENZIMAS DE USO COMUN EN BIOLOGÍA MOLECULAR	54
18- AISLAMIENTO Y PURIFICACIÓN DE FRAGMENTOS DE DNA.	55
18.1- Geles de poliacrilamida	55
18.2- Técnica del Gene-Clean	55
18.3- Técnica de la β -Agarasa	55
19- SECUENCIACIÓN DEL DNA	56
20- TÉCNICAS DE HIBRIDACIÓN DEL DNA Y RNA	56
20.1- Southern-blot	56
20.2- Hibridación de DNA sobre filtros de colonias celulares	57
20.3- Northern-blot	57
24- TÉCNICA DE PRIMER EXTENSION PARA LA DETERMINACIÓN DE LOS SITIOS DE INICIACIÓN DE LA TRANSCRIPCIÓN.	57
25- REACCIÓN EN CADENA DE LA POLIMERASA.	58

III- RESULTADOS 59

Clostridium acetobutylicum 60

1- CARACTERIZACIÓN MOLECULAR DEL GEN ESTRUCTURAL <i>cspA</i> .	61
1.1- Clonación del gen <i>cspA</i>	61
1.2- Análisis de la secuencia de nucleótidos del gen <i>cspA</i>	64
1.3 Análisis de la secuencia de aminoácidos y comparación con otras proteínas del banco de datos	64
1.4- Fraccionamiento subcelular	68
1.5- Expresión del gen <i>cspA</i> en <i>E. coli</i>	69
1.6- Análisis de la presencia del dominio de unión a colina en otras cepas de <i>C. acetobutylicum</i>	73
2- CARACTERIZACIÓN MOLECULAR DE UNA FAMILIA DE GENES QUE CODIFICAN PARA PROTEÍNAS CON DOMINIOS DE UNIÓN A COLINA	75
2.1- Clonación y secuenciación del gen <i>cspB</i>	76
2.2- Clonación y secuenciación del gen <i>cspC</i>	78
2.3- Clonación y secuenciación del gen <i>cspD</i>	81
2.4- Comparación de las secuencias de nucleótidos y aminoácidos de los genes <i>cspA</i> , <i>cspB</i> , <i>cspC</i> y <i>cspD</i>	84
2.5- Utilización de codones de los genes <i>cspA</i> , <i>cspB</i> , <i>cspC</i> y <i>cspD</i>	96
2.6- Estudio de la transcripción de los genes <i>cspA</i> , <i>cspB</i> , <i>cspC</i> y <i>cspD</i>	96
2.7- Localización cromosómica de los genes <i>cspA</i> , <i>cspB</i> , <i>cspC</i> y <i>cspD</i>	98

Streptococcus pneumoniae 101

3- CARACTERIZACIÓN MOLECULAR DEL GEN ESTRUCTURAL <i>pcpA</i> .	102
3.1- Hibridación heteróloga del DNA genómico de <i>S. pneumoniae</i>	102
3.2- Clonación del gen <i>pcpA</i>	104
3.3- Secuenciación del gen <i>pcpA</i>	106
3.3.1- Análisis de la secuencia de nucleótidos del gen <i>pcpA</i>	107
3.3.2- Análisis de la secuencia de aminoácidos	115
3.4.- Comparación de la proteína PepA con otras proteínas del banco de datos.	117
3.5- Caracterización de la unidad de transcripción del gen <i>pcpA</i>	119
3.6- Obtención de mutantes del gen <i>pcpA</i>	119
3.7- Clonación del gen <i>pcpA</i> en neumococo	124
3.8- Localización del gen <i>pcpA</i> en el genoma de <i>S. pneumoniae</i>	124
4- CARACTERIZACIÓN DE UNA NUEVA SECUENCIA DE INSERCIÓN (IS1381) EN <i>S. pneumoniae</i>	128
4.1- Análisis de la secuencia de nucleótidos y aminoácidos: comparación con otras proteínas del banco de datos.	128
4.2- Frecuencia y distribución de la IS1381 en diferentes cepas de <i>S. pneumoniae</i>	129
5- CARACTERIZACIÓN MOLECULAR DE LOS GENES <i>pcpB</i> y <i>pcpC</i>	132

5.1- Clonación de los genes <i>pcpB</i> y <i>pcpC</i>	132
5.2- Análisis de la secuencia de nucleótidos y aminoácidos	135
5.3- Localización de los genes <i>pcpB</i> y <i>pcpC</i> en el genoma de <i>S. pneumoniae</i> . .	140
5.4- Análisis del número de genes codificantes para proteínas con dominios de unión a colina en <i>S. pneumoniae</i>	141

IV- DISCUSIÓN	143
-------------------------	-----

1- CARACTERIZACIÓN MOLECULAR DE LAS PROTEÍNAS Csp DE <i>C. acetobutylicum</i> NCIB 8052	144
2- CARACTERIZACIÓN MOLECULAR DE LA PROTEÍNA PcpA DE <i>S. pneumoniae</i>	149
3- CARACTERIZACIÓN DE UNA NUEVA SECUENCIA DE INSERCIÓN (IS1381) EN <i>S. pneumoniae</i>	157
4- CARACTERIZACIÓN MOLECULAR DE LAS PROTEÍNAS PcpB y PcpC DE <i>S. pneumoniae</i>	162
5- EL DOMINIO DE UNIÓN A COLINA	166

V- CONCLUSIONES	170
---------------------------	-----

VI- BIBLIOGRAFÍA	174
----------------------------	-----

I- INTRODUCCIÓN

1- MICROORGANISMOS QUE CONTIENEN COLINA EN LA ENVOLTURA CELULAR

La colina es un aminoalcohol que forma parte de los fosfolípidos de membrana de las células eucarióticas, siendo poco frecuente en procariotas. Aunque se ha descrito su presencia en los fosfolípidos de membrana de diferentes bacterias (Hagen y cols., 1966; Arondel y cols., 1993), lo más relevante ha sido su descubrimiento, en algunas bacterias Gram-positivas, como componente de los ácidos teicoicos de pared y lipoteicococcos de membrana.

1.1- *Streptococcus pneumoniae*

S. pneumoniae fue el primer microorganismo en el que se describió la presencia de colina como componente de los ácidos teicoicos de pared (Tomasz, 1967). Es una bacteria Gram-positiva, de 0,5-1,25 μm de diámetro, de forma oval con la extremidad distal lanceolada, inmóvil y no forma endosporas. Estas bacterias se presentan en forma diploidea de ahí el nombre de *Diplococcus pneumoniae*, como se le denominó en un principio (Tomasz, 1981). Se aisló por primera vez en el año 1881 por Sternberg (en los Estados Unidos) y Pasteur (en Francia) a partir de muestras de saliva humana.

1.1.1- Importancia desde el punto de vista clínico y biológico

S. pneumoniae es un patógeno humano responsable de enfermedades como la neumonía, meningitis y bacteriemias, provocando altas tasas de mortalidad y morbilidad. Según la Organización Mundial de la Salud (OMS), la neumonía es actualmente la cuarta causa de mortalidad en el mundo y neumococo, en concreto, es el agente causal de entre el 10 y el 25% de los casos (Sahn, 1990). Además de estas infecciones invasivas graves también puede llegar a producir otitis media y sinusitis. Coloniza las vías altas del sistema respiratorio y más concretamente la nasofaringe desde los primeros días de vida y se considera que, virtualmente, cualquier individuo ha estado colonizado en cualquier etapa de su vida. La frecuencia de portadores puede variar de unas comunidades a otras dependiendo de muchos factores (edad, ambiente, estación del año, etc.), y se ha estimado que más del 60% de la población puede ser portador (Austrian, 1986).

En la primera mitad de este siglo, Avery y cols. (1944) establecieron que existían numerosos serotipos de *S. pneumoniae* en función de la estructura polisacáridica de la cápsula. Fue la primera sustancia no proteica que mostraba antigenicidad, de manera que anticuerpos dirigidos contra la cápsula proporcionaban protección contra la virulencia de neumococo. Este hecho fue de gran importancia para el desarrollo de la inmunología. Sin embargo, el estudio del fenómeno de transformación capsular (Griffith, 1928) fue determinante para el descubrimiento del DNA como la base de la información genética, siendo ésta una de las mayores contribuciones al estudio de la biología y medicina, así como la piedra angular del nacimiento de la genética molecular (Avery y cols., 1944).

La reacción de hinchamiento capsular (quellung) junto con la sensibilidad del neumococo frente a las sales biliares (Neufeld, 1900, 1902), son dos de las pruebas más utilizadas en el diagnóstico clínico. Neumococo fue posiblemente la primera especie bacteriana en la que se detectó resistencia *in vivo* a un agente antimicrobiano, la optoquina (derivado de la quinina) (Morgenroth y Kaufmann, 1912). Hoy en día la resistencia de esta bacteria a agentes antimicrobianos se ha incrementado notablemente, acrecentándose el peligro potencial de la misma. También ha sido neumococo la primera bacteria en la que se ha estudiado la naturaleza enzimática de la lisis bacteriana (Dubos, 1937) y de la cual se ha clonado el gen de la primera autolisina bacteriana (García y cols., 1985).

1.1.2- Aspectos estructurales de las envolturas de neumococo

1.1.2.1- La pared celular

La pared celular de neumococo está compuesta del peptidoglicano y de los ácidos teicoicos a él asociados. Su espesor oscila entre los 14-16 nm, y está estructurada en 6-8 capas de peptidoglicano de 2 nm de espesor cada una (Tomasz, 1981). El peptidoglicano bacteriano es un heteropolímero formado por cadenas polisacáridicas interconectadas por puentes peptídicos. La fracción polisacáridica, está constituida habitualmente por la alternancia de residuos de N-acetilglucosamina (NAG, 2-*N*-acetilamino-2-desoxi-D-glucosa) y ácido N-acetilmurámico (NAM, 2-*N*-acetilamino-3-(*O*)-(D-1-carboxietil)-2-desoxi-D-glucosa), unidos entre sí mediante enlaces β -1,4. Al grupo carboxilo de los restos de NAM se encuentran unidas las cadenas peptídicas a través de su grupo amino terminal. La estructura del componente peptídico del peptidoglicano de neumococo ha sido estudiado

por García-Bustos y cols. (1987, 1990), gracias al empleo de cromatografía líquida de alta presión (Fig. 1). El principal péptido monomérico es el tripéptido (L-Ala-D-iGln-L-Lys) en lugar de un tetrapéptido (L-Ala-D-iGln-L-Lys-D-Ala), lo que puede indicar la presencia de D,D y D,L-carboxipeptidasas. Existen además dos tipos de puentes interpeptídicos: el más abundante es un puente cruzado L-Ala-L-Ser entre la D-Ala en posición 4 de la unidad peptídica donadora y la L-Lys de la posición 3 de la aceptora; el segundo tipo de puente es el enlace directo entre estos dos aminoácidos, D-Ala(4) y L-Lys(3). Además de estos aminoácidos, también se han encontrado, aunque en proporción minoritaria, Gly y Asp, los cuales se supone que forman parte de puentes interpeptídicos (Fischer y Tomasz, 1985).

Respecto a los ácidos teicoicos, podemos decir en general que forman parte de la pared celular de bacterias Gram-positivas y pueden constituir hasta el 10% del peso seco de la célula. Son polímeros de ribitol ó glicerol fosfato, en donde los monómeros se unen mediante enlaces fosfodiéster. En el caso de neumococo, se les ha denominado también como polisacárido C y su naturaleza antigénica se atribuye a los restos de N-acetil galactosamina fosfato (Tomasz, 1981). Mediante técnicas inmunocitoquímicas se ha visto que se localiza no sólo en la superficie de los neumococos no encapsulados sino en las caras internas y externas de paredes celulares aisladas (Sørensen y cols, 1988). Constituyen el 40-50% del peso seco de la pared, estando unidos covalentemente al peptidoglicano a través de enlaces tipo fosfodiéster. Durante la síntesis de la pared celular, los ácidos teicoicos se van incorporando al peptidoglicano insoluble y parece ser que el entrecruzamiento del peptidoglicano actuaría como señal para designar los restos de ácido murámico que pueden incorporar cadenas de ácidos teicoicos (Fischer y Tomasz, 1984). Se estima que debe existir, por término medio, una cadena de ácido teicoico con 8 subunidades por cada unidad de tetradisacárido tetrapéptido (García-Bustos y Tomasz, 1987). En cuanto a su función, los ácidos teicoicos son capaces de modular la actividad de las enzimas líticas, de activar la ruta del complemento y pueden actuar como receptores de bacteriófagos (López y cols, 1982b).

Se ha propuesto que los ácidos teicoicos de *S. pneumoniae* están compuestos de unidades repetidas de glucosa-tridesoxidiaminohexosa-N-acetilgalactosamina-N-acetilgalactosamina-ribitol fosfato. Por cada unidad de ácido teicoico hay dos residuos de fosforilcolina situados en la posición O(6) de los residuos de N-acetilgalactosamina

(Fischer, 1993) (Fig. 1). El 85% de la colina macromolecular se encuentra unida al ácido teicoico (Brundish y Baddiley, 1968; Mosser y Tomasz, 1970).

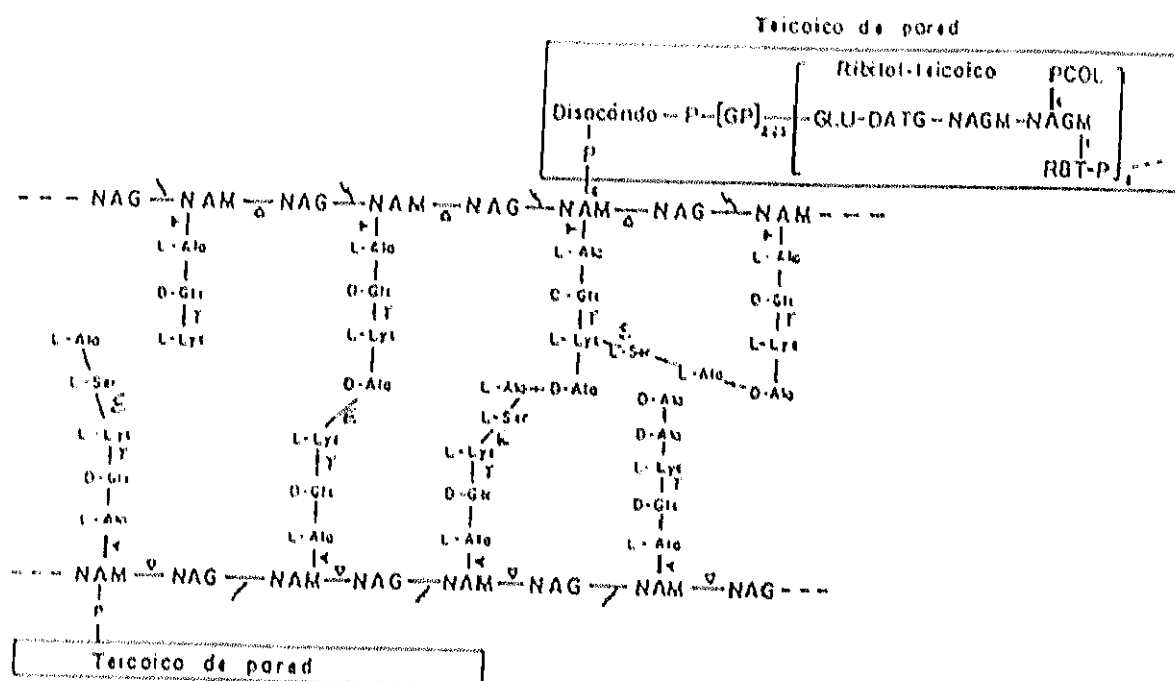


Figura 1- Modelo de la estructura de pared celular de *S. pneumoniae*. Las cadenas glicánicas se representan con las letras NAG-NAM, que representan los residuos alternantes de *N*-acetilglucosamina (NAG) y *N*-acetilmurámico (NAM). e significa que el enlace peptídico se realiza con el grupo ε-amino de la L-lisina. El ácido teicoico de pared se simboliza con un rectángulo. P, grupo fosfato; GIP, glicerol fosfato; GLU, glucosa; DATG, diacetamidotridesoxigalactosa; NAGM, *N*-acetilgalactosamina; PCOL, fosforilcolina; RBT-P, ribitolfosfato; Glx, isoglutamina. Se estima que hay, por término medio, 16 moléculas de PCOL en cada molécula de ácido teicoico. Los números 1 y 6 representan las posiciones de los átomos de carbono que participan en los enlaces que se indican. Los enlaces sensibles a la acción de las enzimas que hidrolizan la pared celular de neumococo se muestran con triángulos negros (sensibles a las amidasas), con triángulos blancos (sensibles a las muramidasa o lisozimas) y con flechas en zig-zag (sensibles a glucosaminidasas).

S. pneumoniae presenta la peculiaridad de poseer un requerimiento nutricional absoluto por la colina, pues la eliminación de la misma del medio de cultivo provoca un cese del crecimiento (Rane y Subbarow, 1940). Se ha visto que su ausencia inhibe la síntesis y liberación tanto del peptidoglicano como del ácido teicoico (Fischer y Tomasz, 1984). Este requerimiento nutricional puede satisfacerse utilizando análogos estructurales (Badger, 1944; Sanz y cols, 1988), como la etanolamina, la cual neumococo es incapaz de metilar *in vivo* (Tomasz, 1968), dando lugar a alteraciones fisiológicas importantes como la no separación de las células hijas, la incapacidad de lisarse al final de la fase estacionaria de crecimiento, la resistencia a la lisis inducida por desoxicolato y la incapacidad de transformarse (Tomasz, 1968). Posteriormente se observó también la tolerancia a antibióticos β -lactámicos (Tomasz, 1970) y la no infección por fagos.

Las propiedades anómalas mostradas por las células EA, desaparecen mediante la adición de colina a los cultivos de neumococo que contienen etanolamina. Las nuevas unidades de ácido teicoico conteniendo colina se unen a los precursores del peptidoglicano sintetizado simultáneamente (Tomasz y cols, 1975), formando complejos que se incorporan en la pared, exclusivamente en la región ecuatorial de la misma (Briles y Tomasz, 1970; López y cols, 1982b). Durante el crecimiento celular, se segregan de forma simétrica y pasan intactos a las células hijas. De esta manera todas las alteraciones fisiológicas mencionadas anteriormente desaparecen.

Las células cuya pared contienen EA poseen una enzima autolítica denominada LytA, en una conformación que posee una actividad muy reducida (forma E) pero que puede activarse (convertirse) a la forma catalíticamente activa (forma C) tras la preincubación de la misma a baja temperatura con paredes que contienen colina (Tomasz y Westphal, 1971).

1.1.2.2- La membrana plasmática

En cortes ultrafinos observados al microscopio electrónico la membrana plasmática de neumococo posee la típica estructura trilamelar, con un espesor aproximado de 9 nm (Fig. 2) y con una composición lipídica en la que destacan los glicolípidos, el fosfatidilglicerol, la cardiolipina y los lípidos neutros (Horne y cols., 1977; Tomasz., 1981; Sørensen y cols., 1988).

Un componente importante de la membrana es el antígeno de Forssman, molécula anfipática que presenta una estructura idéntica a la del ácido teicoico en cuanto a unidades repetidas se refiere (Behr y cols., 1992; Fischer y cols., 1993), y que lleva material lipídico unido covalentemente (se denomina también ácido lipoteicoico o LTA). Se encuentra uniformemente distribuido en la membrana plasmática con la fracción lipídica anclada en la bicapa lipídica, quedando el resto de la molécula expuesto en la superficie (Sørensen y cols., 1988). El 15-20% de la colina de la envuelta celular forma parte de este antígeno, también denominado antígeno F

Otras proteínas interesantes desde el punto de vista funcional y que se localizan en la membrana son las proteínas que unen penicilina ó PBPs, el receptor del factor de competencia (Tomasz, 1968), las proteínas implicadas en la unión del DNA exógeno durante el proceso de transformación genética (Tomasz, 1981) y la endonucleasa encargada de digerir una de las dos hebras de DNA transformante, requisito indispensable para la entrada de DNA en la célula.

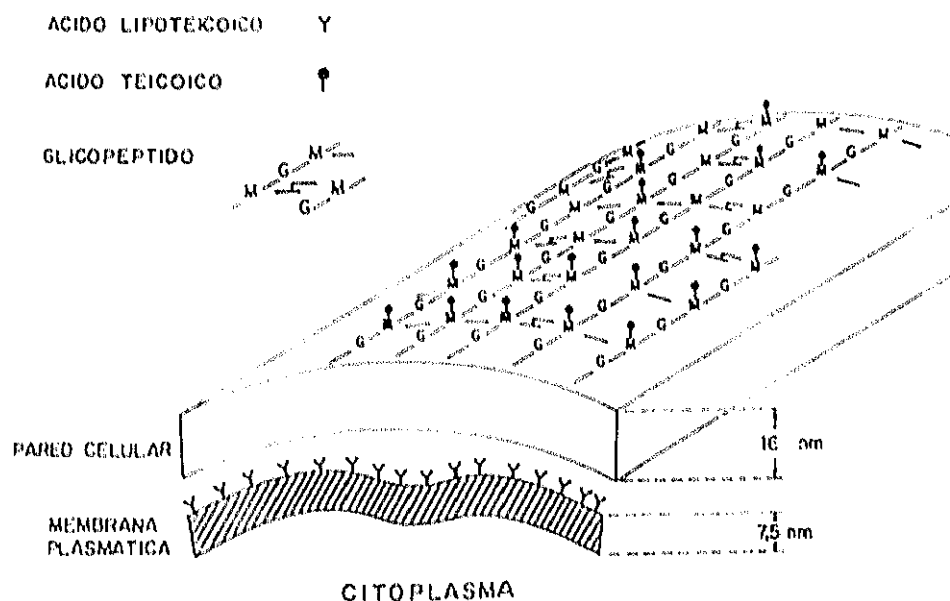


Figura 2- Representación esquemática de la membrana plasmática y la pared celular de *S. pneumoniae*. Abreviaturas: G, N-acetilglucosamina; M, ácido N-acetilmurámico. Tomado de Tomasz (1981).

1.1.2.3- La cápsula

Durante muchos años la cápsula fue considerada responsable de la virulencia producida por neumococo. Esto se basaba en la observación de que todos los aislados clínicos eran capsulados (smooth), mientras que los derivados no capsulados (rough) de dichas cepas, eran completamente avirulentos. Aún no se conoce bien el mecanismo por el cual la cápsula provoca virulencia, aunque sí se sabe que presenta propiedades antifagocitarias (Paton y cols., 1993). El espesor de la cápsula varía entre 200 y 400 nm dependiendo de las estirpes. Aunque no se conoce con precisión si la cápsula se encuentra unida a la pared celular por algún tipo de interacción, se ha descrito que en algunos tipos capsulares puede unirse al peptidoglicano mediante enlaces covalentes (Sørensen y Blom, 1992; Sørensen y cols., 1990). La cápsula es de naturaleza polisacáridica y también recibe el nombre de sustancia SSS (specific soluble substance). Se conocen 84 serotipos de neumococo, que se diferencian entre sí, tanto por los monosacáridos que constituyen la cápsula como por el tipo de enlace entre los mismos. Los 84 serotipos se agrupan en 46 grupos, si bien hay serotipos con antígenos comunes que pertenecen a distintos grupos (Lund y Henriksen, 1978).

1.2- *Clostridium acetobutylicum*

1.2.1- Propiedades de *C. acetobutylicum* y clasificación

El género *Clostridium* abarca unas 100 especies de microorganismos anaerobios, Gram-positivos, formadores de endosporas, pertenecientes a líneas filogenéticas separadas. Se incluyen especies psicrófilas, mesófilas y termófilas, mostrando una amplia diversidad en cuanto a la morfología, el metabolismo y los requerimientos nutricionales, variando su contenido en G+C entre un 24-54 mol %. A finales del siglo pasado todas las especies aisladas del género *Clostridium* se clasificaban en patógenas y no patógenas. El primer grupo se caracterizaba por la producción de potentes toxinas, mientras que el segundo presentaba distintas propiedades, como la capacidad de fijar nitrógeno, producir solventes, secretar enzimas responsables de la degradación de polímeros, etc. El papel más importante que desempeñan en la naturaleza es la degradación de material orgánico, polímeros tales como almidón, celulosa, hemicelulosa y proteínas. Muchas especies secretan enzimas como

amilasas, celulasas, proteasas y colagenasas, algunas de las cuales son de gran interés biotecnológico (Saha y cols., 1989). Las especies de *Clostridium* mejor estudiadas son las patógenas para el hombre (*C. botulinum*, *C. difficile*, *C. perfringens*), *C. thermocellum* por su capacidad de degradar la celulosa a elevadas temperaturas y bacterias como *C. acetobutylicum*, *C. beijerinckii*, *C. tetanomorphum* y *C. aurantibutyricum* (Rodgers, 1986) responsables de la fermentación acetobutílica. *C. acetobutylicum* es hasta ahora el microorganismo mejor caracterizado y el mejor productor natural de butanol.

Fue Weizman en el año 1916 quien desarrolló la explotación a gran escala de la fermentación acetobutílica, estableciéndose la primera planta de producción en Inglaterra. A partir de la primera estirpe productora se han venido utilizando otras estirpes productoras que se asemejan a la de Weizman en cuanto a la producción de acetona y butanol quedando tipificadas como la misma especie. En la tabla 1 se muestran las estirpes mejor caracterizadas de *C. acetobutylicum*, siendo la estirpe ATCC 824 la cepa tipo. A pesar de que todas estas estirpes son capaces de llevar a cabo la fermentación acetobutílica, se ha visto que presentan importantes diferencias fenotípicas y genotípicas (Wilkinson y cols., 1995). Dada la heterogeneidad del grupo era interesante conocer cuales son clasificables como *C. acetobutylicum* y cuales no.

Tabla 1- Estirpes de *Clostridium acetobutylicum*

Estirpe	Referencia
ATCC 824	Croux y cols., 1992b
NCIB 8052	García y cols., 1988b
P262	Woolley y Morris, 1990
DSM 1731	Woolley y Morris, 1990
NI-4 ^a	Hongo, 1960

^a Esta cepa fue antiguamente denominada *C. scharoperbutylacetonicum*

En los trabajos realizados por Woolley y cols. (1990), se llegó a la conclusión de que al menos cuatro de las cinco estirpes mencionadas anteriormente podrían subdividirse en dos grupos claramente diferenciados en cuanto a su crecimiento, características

fisiológicas, bioquímicas y de fermentación, sumándose la poca o nada de homología entre sus DNAs. Las cepas NCIB 8052 y P262 conformarían un grupo y la cepa DSM 1731 y ATCC 824 conformaría el otro. A su vez, los estudios de espectrofotometría de masas por pirolisis (Wilkinson y cols., 1995) así como el estudio de los sustratos fermentables por dichas cepas (Woolley y cols., 1990) marcó ciertas diferencias entre las estirpes NCIB 8052 y la P262, quedando incluso esta última más relacionada con la cepa NI-4081 (derivada de la NI-4) (Wilkinson y cols., 1995). Con respecto a las diferencias fenotípicas, parecen estar relacionadas con la composición de los ácidos teicoicos de pared y la actividad de sus enzimas liticas como veremos en apartados posteriores (Podvin y cols., 1988; Croux y Garcia, 1991). En relación a las diferencias genotípicas, Wilkinson y cols. (1993) hicieron un estudio del tamaño del genoma de dichas estirpes mediante la técnica de electroforesis en campo pulsado, distinguiéndose claramente tres grupos. Las estirpes NCIB 8052 y NI-4081 tenían un tamaño del DNA genómico de 6,5 Mpb con un perfil de restricción idéntico. Las estirpes ATCC 824 y DSM 1731 presentaban unos tamaños de 4 y 3,5 Mpb respectivamente, siendo los perfiles de restricción del DNA muy similares. La estirpe P262 presentaba, sin embargo, un tamaño del DNA genómico de 2,85 Mpb. A estos estudios hay que sumar las comparaciones de las secuencias de los rRNA 16S que confirmaron la estrecha relación entre las estirpes ATCC 824 y DSM 1731, al presentar idénticas secuencias. La estirpe NCIB 8052 mostraba un 7% de divergencia frente a las estirpes anteriores, siendo su secuencia 100% idéntica con el rRNA 16S de *C. beijerinckii*, indicando que estas estirpes representarían la misma especie. La estirpe P262 se encuentra bastante relacionada con las anteriores aunque mostraba cierta divergencia. Otras diferencias interesantes a destacar entre estas estirpes son, por ejemplo, la presencia de un sistema de restricción activo en la estirpe ATCC 824 ó la capacidad de movilizar plásmidos desde *E. coli* a la estirpe NCIB 8052 pero no a la ATCC 824 ó P262.

Todos estos datos venían a reafirmar aún más la heterogeneidad de todas estas estirpes, las cuales podrían no constituir la misma especie. Estas investigaciones, sugirieron a Wilkinson y cols. (1995) que el nombre de *C. acetobutylicum* debería restringirse exclusivamente a aquellas estirpes genéticamente relacionadas con la estirpe tipo ATCC 824, considerando que la estirpe NCIB 8052 debería denominarse *C. beijerinckii*.

1.2.2- Importancia industrial: la fermentación acetobutílica (AB)

C. acetobutylicum ha sido uno de los microorganismos más ampliamente utilizados para el desarrollo a gran escala de la fermentación acetona-butanol (Jones y Woods, 1986). Durante la primera guerra mundial, el coste de producción se invirtió en la producción de acetona para la elaboración de municiones, pues no existía un mercado para el butanol, el cual constituía el 60% de la producción de solventes. Después de la guerra, con el desarrollo de la industria automovilística, comenzó la demanda de butanol, incrementándose el rendimiento del proceso de fermentación mediante la utilización de cepas más productoras de dicho solvente. Al mismo tiempo se hicieron grandes esfuerzos para trabajar en una planta de fermentación integrada, donde los solventes se obtenían por destilación y los gases emanados se separaban y se recogían, utilizándose el CO₂ como gas comprimido para la producción de nieve carbónica y el H₂ para la producción de metanol o la hidrogenación de aceites en la producción de margarina. Los residuos sólidos de la fermentación se vio que eran ricos en riboflavina, utilizándose como complemento en los alimentos para animales. A pesar de que la fermentación acetona-butanol sobrevivió durante la segunda guerra mundial, no podía competir con la producción química de solventes a partir del petróleo por las siguientes razones: i) la necesidad de una fuente de carbohidratos como sustrato del proceso fermentativo, cuyo precio fluctúa según la cosecha, la demanda del mercado y la intervención del gobierno; ii) los problemas derivados del coste adicional que supone la eliminación o reciclaje de los productos de desecho; iii) el alto coste que supone el proceso de destilación para la separación de los solventes orgánicos comparado con el proceso de condensación utilizado en las plantas petroquímicas, donde la mayor parte de los procesos tienen lugar en fase gaseosa. Para hacer competitivo el proceso de fermentación es necesario mejorar, no sólo el proceso microbiológico sino también la ingeniería bioquímica.

En la figura 3 se resumen las principales rutas metabólicas del proceso de fermentación. Durante el proceso de fermentación podemos distinguir dos fases fisiológicas diferentes. La fase inicial de la fermentación está asociada con el crecimiento celular exponencial, es la denominada fase acidogénica y se caracteriza por la producción de ácido acético, butírico, hidrógeno, dióxido de carbono así como lactato en algunas ocasiones, dando todo ello lugar a una disminución del pH del medio. Durante la segunda etapa de la fermentación, denominada solventogénica, la tasa de crecimiento disminuye y se produce un

cambio de la producción de ácidos hacia la producción de solventes (etanol, acetona y butanol). Este proceso va acompañado de una disminución del hidrógeno molecular y un aumento en la producción de dióxido de carbono. Por otro lado, los ácidos producidos en la etapa anterior son reasimilados, incrementándose el pH del medio de cultivo (Jones y Woods, 1986). Se ha visto que la producción de acetona durante la fase solventogénica está directamente ligada a la reutilización de acetato y butirato generados durante la fase inicial de la fermentación a través de la actividad de la acetil CoA transferasa, lo que implica que no puede haber un consumo de ácido sin la producción equimolecular de acetona (Hartmanis y cols., 1984). Esto indica, que en un cultivo discontinuo no se podría obtener un buen rendimiento en la producción de butanol sin el consiguiente consumo de ácidos que llevan a la producción de acetona. Sin embargo, experimentos recientes han demostrado que la producción de acetona no necesita estar directamente emparejada con la formación de butanol, puesto que la adición de acetato en condiciones de exceso de glucosa resulta en un marcado incremento en la producción de acetona, pero no en la producción de butanol (Matta-El-Amouri y cols., 1985). La reutilización de ácidos parece más bien significar un mecanismo de detoxificación de las células. La etapa final de la fermentación tiene lugar cuando la concentración de solventes llega hasta niveles inhibitorios, cesando el metabolismo celular.

Una de las mayores limitaciones de la fermentación AB es la baja concentración de solventes al final del proceso de fermentación. El proceso de fermentación discontinuo se detiene cuando la concentración de solventes llega a 18-22 g/l (Jones y Woods, 1986). La baja concentración de solventes al final del proceso de fermentación implica que la recuperación de los mismos por destilación es un proceso energéticamente desfavorable al que hay que añadir todo el coste de producción. El efecto tóxico del butanol se manifiesta cuando se encuentra presente en el medio en concentraciones de 12 y 16 g/l. El mecanismo de la toxicidad del butanol no está todavía muy bien estudiado pero parece ser que es debido a una alteración en la membrana celular (Ingram, 1986; Hutkins y Kashket, 1986). Ésta parece ser también la causa de la toxicidad del etanol en bacterias mesófilas y levaduras. La potencia de estos alcoholes como inhibidores parece ir ligada a su solubilidad en lípidos. Tanto el etanol como el butanol son moléculas anfipáticas, una situación que les permite unirse a la región hidrofóbica de la membrana, incrementando por otro lado, la polaridad de la misma. Esto trae consigo un doble efecto: una disminución de la eficiencia

de la membrana como barrera de permeabilidad selectiva, provocando incluso roturas en la misma, y una alteración en la organización de los fosfolípidos y proteínas de membrana. Aunque el modo de acción del butanol y del etanol presenten similitudes, sus mecanismos específicos parecen ser diferentes (Ingram, 1986). La introducción del butanol en la membrana parece incrementar la fluidez de la misma favoreciendo su rotura. Se ha visto que la adición de butanol provoca una disminución de la actividad ATPásica ligada a la membrana, provocando un colapso en el gradiente de pH debido a la alteración de la bomba de protones (Moreira y cols., 1981; Bowles y Ellefson, 1985; Gottwald y Gottschalk, 1985). La adición de butanol provoca también una disminución en el nivel del ATP intracelular pero no existe ninguna evidencia directa de la inhibición de la glicolisis por el butanol (Ingram, 1986).

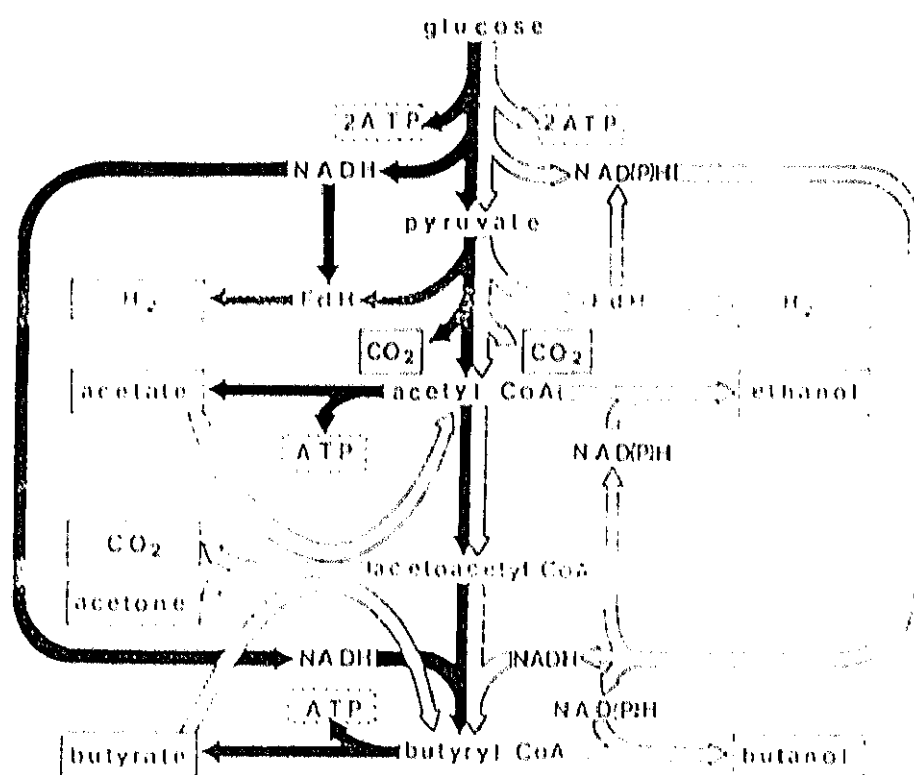


Figura 3- Rutas bioquímicas de la fermentación acetobutílica. Las flechas negras corresponden a la fase acidogénica y las blancas a la solventogénica. Tomado de Jones y Woods, 1988.

La toxicidad del butanol se ha visto también ligada a la autólisis de las células productoras de solventes mediante la inducción y liberación de autolisina. Un mutante pleiotrópico deficiente en la actividad autolítica, fue más resistente a la degeneración celular en presencia de butanol, produciendo además mayores concentraciones de solventes y butanol que las estirpes parentales. (Van der Westhuizen y cols., 1982). De todo esto resulta evidente que una mayor tolerancia de las células al butanol supondría una mayor producción de solventes.

La manipulación de las células productoras de solventes para evitar la degeneración celular, así como la resistencia a autolisinas, bacteriocinas y bacteriófagos, podría mejorar la estabilidad de las estirpes utilizadas para la producción de solventes en procesos de fermentación continuos o discontinuos.

1.2.3- Aspectos estructurales de la pared de *C. acetobutylicum*

Sóloamente algunos trabajos y relativamente antiguos, son disponibles sobre la estructura del peptidoglicano del género *Clostridium*. Takumi y Kawata (1970) y Cummins y Johnson (1971) determinaron, para diferentes especies de dicho género, la presencia de m-DMP, alanina y glutámico como los aminoácidos presentes en el peptidoglicano, a una concentración coherente con la formación de puentes directos entre las unidades tetrapeptídicas. Este tipo de estructura fue la que establecieron Ogata y cols. (1975) para el caso de *C. acetobutylicum* NI-4. En el caso de la cepa tipo ATCC 824, excepto la presencia de m-DMP, no se tiene ninguna información adicional sobre la estructura de su peptidoglicano.

No se ha descrito en la literatura ninguna purificación ni caracterización de los ácidos teicoicos o lipoteicoicos, salvo la presencia colina en los ácidos teicoicos de pared de *C. acetobutylicum* NI-4 (Podvin y cols., 1988). La identificación de una enzima lítica de pared en *C. acetobutylicum* NCIB 8052 capaz de degradar paredes de neumococo conteniendo colina, sugiere la presencia de este aminoalcohol en las envueltas celulares de dicho microorganismo (García y cols., 1988).

1.3- Otros microorganismos

Aunque sólo se ha descrito la presencia de proteínas con dominios de unión a colina en *S. pneumoniae*, *S. oralis* y en algunas estirpes de *C. acetobutylicum*, la colina es un

componente estructural de la pared de otros estreptococos (Kilpper-Bälz y cols., 1985) (Tabla 2).

Tabla 2- Estirpes de estreptococos que contienen colina en la pared celular

Microorganismo	Estirpe
<i>S. constellatus</i>	DSM 20575 ^T
<i>S. oralis</i>	NCTC 11427 ^T
<i>S. viridans</i> I	Kiel 50493
<i>S. viridans</i> II	Kiel 60973
<i>S. viridans</i> IV	Kiel 42841
<i>Streptococcus</i> sp. group O	Kiel C119/48
<i>Streptococcus</i> sp.	
" <i>S. mitior</i> "- <i>S. sanguis</i>	S229 ⁸ , S410 ⁸ , S426 ⁸ S438 ⁸ , S440 ⁸ , S444 ⁸ S447 ⁸ , S526 ⁸ , S662 ⁸
<i>S. mitis</i>	NCTC 12261

Ronda y cols. (1988) estudiaron la expresión del gen *lytA* que codifica para la amidasa de neumococo dependiente de colina en *S. oralis* aprovechando la presencia de colina en la pared de este microorganismo. La expresión de dicho gen en *S. oralis* provocó su lisis aunque sólo tras la inducción con detergentes. Extractos de *S. oralis* fueron utilizados para medir actividad lítica sobre paredes de neumococo, detectándose una pequeña actividad sobre las mismas que se caracterizó posteriormente como una fosforilcolín esterasa (Ronda y cols., 1991). Es posible que esta proteína posea un dominio de unión a colina que permita el reconocimiento de las paredes de neumococo al igual que la amidasa LytA es capaz de reconocer las de *S. oralis*.

2- EL DOMINIO DE UNIÓN A COLINA: PROTEÍNAS CON CAPACIDAD DE UNIÓN A COLINA

2.1- Autolisinas de neumococo

Las autolisinas son enzimas que hidrolizan uniones específicas de los peptidoglicanos, solubilizando las paredes bacterianas con la consiguiente lisis de la bacteria, de ahí que antiguamente se denominasen "suicidasas". La presencia de enzimas autolíticas en bacterias tanto Gram-positivas como Gram-negativas es un fenómeno casi general, aunque sólo en algunos casos se conoce su mecanismo de acción. También se ha descrito su presencia en hongos y levaduras. Atendiendo a la especificidad del enlace hidrolizado existen cuatro tipos de enzimas líticas (Fig. 1) (Rogers. 1974):

i) Glicosidasas, que hidrolizan la cadena glicánica del peptidoglicano. Entre ellas se encuentran las endo *N*-acetilglucosaminidasas, que liberan fragmentos con residuos de NAM en el extremo reductor, y las endo-*N*-acetilmuramididasas, que liberan fragmentos con residuos de NAG en el extremo reductor.

ii) *N*-acetil muramil-L-alanil amididasas, que hidrolizan el enlace amida entre la L-alanina y el NAM.

iii) Endopeptidasas, que rompen la cadena peptídica del peptidoglicano, y de las cuales se han descrito varios ejemplos según el enlace roto dentro del péptido.

iv) Transglicosilasas, las cuales liberan grupos no reductores de NAM.

Las autolisinas más ampliamente distribuidas son las amididasas y las glucosaminidasas, que en la mayoría de los casos se presentan simultáneamente en el mismo organismo, como ocurre en el caso de *S. pneumoniae*. Las enzimas líticas de este microorganismo requieren para su actividad de la presencia de colina en los ácidos telcoicos de pared como se podrá deducir de lo expuesto a continuación.

2.1.1- La amidasa LytA

Es la enzima lítica mayoritaria presente en neumococo. Fue parcialmente caracterizada por Howard y Gooder (1974) como una *N*-acetil-muramil-L-alanil-amidasa. Un año después fue purificada a homogeneidad electroforética (Höltje y Tomasz, 1976). Presenta un peso molecular de 36.000, que coincide con el deducido de la secuencia de aminoácidos (García y cols., 1986b). El gen se ha clonado y expresado en *E. coli* como la forma inactiva o forma E (Tomasz y Westphal, 1971), siendo ésta por tanto el transcrito primario del gen (García y cols., 1985). La forma E inactiva puede transformarse en la forma C activa tras la incubación a baja temperatura con paredes que contienen colina o simplemente con colina. La actividad de la amidasa es totalmente dependiente de la presencia de los residuos de colina en los ácidos teicoicos de pared. Tanto la eliminación de los ácidos teicoicos de pared como la utilización de paredes celulares en las cuales los residuos de colina se han sustituido biosintéticamente por etanolamina, no son hidrolizados por la enzima (Höltje y Tomasz, 1975b; Mosser y Tomasz, 1970). La conversión de las paredes celulares EA a paredes con colina, mediante metilación química *in vitro*, hace a éstas sensibles a la acción de la autolisina. Puesto que la única diferencia entre las paredes EA y las paredes con colina es el aminoalcohol, la actividad hidrolítica de la autolisina requiere la interacción con los residuos de colina de los ácidos teicoicos (Höltje y Tomasz, 1975b). Es muy probable que la autolisina se una a su sustrato a través de la colina en un paso previo a la hidrólisis del enlace amida (Giudicelli y Tomasz, 1984). El principal papel de los residuos de colina es actuar de ligando de adsorción facilitando el ataque de la enzima soluble a su sustrato insoluble (Giudicelli y Tomasz, 1984). La dependencia de LytA por la colina, así como la inhibición por la misma, desaparece en presencia de un peptidoglicano soluble, de lo que se deduce que el requerimiento por la colina va unido a la naturaleza insoluble del sustrato (García-Bustos y Tomasz, 1987). Además, el mecanismo de degradación de la pared celular no es el mismo *in vivo* que *in vitro* (García Bustos y Tomasz, 1989). La diferencia no estriba en la preferencia por un tipo u otro de unidad peptídica sino en el grado de degradación. En la hidrólisis *in vivo*, más del 94% de las unidades peptídicas se encuentran unidas al complejo azúcar-ácido teicoico, lo que implica que la hidrólisis debe tener lugar mediante la ruptura de enlaces amida estratégicamente situados entre las distintas capas que constituyen el peptidoglicano, permaneciendo intactas las uniones interpeptídicas de cada lámina.

La función de la amidasa, ha podido ser estudiada gracias a la obtención de mutantes *Lyt⁻*. El primer mutante autolisina negativo (cepa *cwl-1*) fue aislado por Lacks (1970). La mutación presente en la estirpe *cwl-1* representó el primer caso de mutación del gen estructural de una autolisina bacteriana. La alteración presente en el gen *lytA* corresponde a la transición de CG a TA, provocando un cambio de una glicina por un resto de ácido glutámico y alterando el perfil hidropático de la proteína (García y cols., 1986a). Se caracterizó otra mutación en el gen *lytA* que se denominó *lyt-32* (cepa M32) y que consiste en una mutación sencilla por delección de un par GC, apareciendo dos codones de terminación. Éste fue el primer mutante descrito defectivo en la actividad amidasa (López y cols., 1986). Finalmente se aisló y caracterizó el primer mutante delecionado en el gen *lytA* y que se denominó M31 (Sánchez-Puelles y cols., 1986a,b). El aislamiento de esta cepa permitió el estudio del papel biológico que desempeña la principal autolisina de neumococo. Este mutante crece a 37°C formando pequeñas cadenas, no se autolisa al final de la fase estacionaria de crecimiento, ni en presencia de detergentes y exhibe una respuesta tolerante frente los antibióticos β -lactámicos. Todas estas características revierten cuando se clona y expresa dicho gen en el mutante (Ronda y cols., 1987). De la existencia de este mutante se dedujo que la amidasa *LytA* no era un enzima indispensable para adquirir el estado de competencia ni para la multiplicación celular. Se ha sugerido también que esta enzima participa como uno de los factores desencadenantes de la virulencia (Berry y cols., 1989).

2.1.1.1- Proceso de conversión

Se sugirió que el proceso de activación de la autolisina *LytA* (conversión) implicaba algunos cambios estructurales o conformacionales de la forma E, esenciales para el proceso catalítico (Höltje y Tomasz, 1976). Las células EA de neumococo contienen una forma inactiva de la amidasa o forma E (Tomasz y Westphal, 1971). La forma E de la amidasa puede ser convertida *in vitro* a la forma activa ó forma C tras la incubación con paredes celulares que contienen colina (Tomasz y Westphal, 1971) o en presencia de 2% de colina (Briese y Hakenbech, 1985). Dependiendo del tipo de conversión, parte de los ácidos teicoicos conteniendo colina o moléculas de colina libre, se unen de manera no covalente a la forma C activa dando lugar a una forma de alto M_r (500.000) ó de bajo M_r (36.000) según el caso. Hay que señalar que la forma activa de alto M_r no es debida a la formación de agregados de la enzima sino el resultado de la unión de la enzima a los ácidos teicoicos de la

pared celular. Por otro lado, una prolongada diálisis de la forma C de bajo M_r , resulta en la reconversión a la forma inactiva E indicando que la actividad de la forma C es dependiente de las moléculas de colina unidas a la enzima (Höltje y Tomasz, 1975b; Briesse y Hakenbeck., 1985). Los trabajos realizados por Sánchez-Puelles y cols. (1987), demostraron que la región C-terminal de la proteína está implicada en el proceso de activación de la enzima y aunque la región C-terminal no es esencial para la actividad catalítica si se requiere un C-terminal intacto para conseguir una funcionalidad completa. Estos mismos autores han visto que los 11 últimos aminoácidos de LytA son esenciales para el proceso de conversión pues la delección de los mismos mediante manipulación genética, da lugar a una forma de la amidasa que no requiere del proceso de conversión. Estos últimos 11 aminoácidos corresponden a una cola hidrofóbica que podría interaccionar con otras regiones hidrofóbicas de la molécula o quizás con la membrana celular, plegando la enzima de tal manera que el centro activo permanezca escondido. En la conversión, la unión de la forma E de la amidasa a la colina, modificaría el plegamiento anterior presentando el centro activo al sustrato.

Un gran número de compuestos con una estructura química similar a la de la colina son capaces de convertir la forma E de la amidasa a la forma C. De la utilización de dichos compuestos se ha deducido que las alquilaminas terciarias son las estructuras mínimas que se requieren para la activación catalítica *in vitro* de la forma E de la amidasa. Por otro lado, la presencia de sustituyente polares en estos análogos bloquea el proceso de activación, pudiéndose especular sobre la importancia de la interacción hidrofóbica entre los derivados de colina y los dominios hidrofóbicos de la forma E de la amidasa para la activación catalítica de la enzima (Sanz y cols., 1988). Algunos de estos compuestos se han utilizado con éxito para la rápida purificación de la amidasa así como de otras enzimas líticas dependientes de colina (García y cols. 1987; Sanz y cols. 1988).

2.1.1.2- Inhibidores de la actividad catalítica

Se han descrito al menos dos inhibidores de la actividad lítica de la amidasa: el ácido lipoteicoico y la colina. La inhibición por el LTA es de tipo no competitivo siendo el proceso reversible tras la adición de desoxicolato sódico (DOC) (Höltje y Tomasz, 1975a). La interacción entre el LTA y la amidasa se utilizó para la purificación de la enzima mediante cromatografía de afinidad en columnas de LTA-Sefarosa (Briesse y Hakenbeck,

1985). La proteína se eluía con colina lo que implicaba que la unión enzima-LTA se llevaba a cabo a través de los residuos de colina. Por otro lado, la enzima purificada resultó ser la forma C activa en contra de lo que se esperaba dado el carácter inhibitorio del LTA. Se sabe que la presencia de lípidos en el LTA conduce a la formación de micelas en solución acuosa las cuales impedirían el acceso de la enzima a su sustrato actuando, por tanto, como inhibidores de la actividad litica. La adición de lipasa o de DOC impedirían la formación de agregados micelares lo que explicaría la desinhibición de la enzima litica por LTA tras la adición de ambos compuestos. Interacciones similares parecen tener lugar en las fracciones de membrana procedentes de células multiplicadas en presencia de colina, lo que sugiere que la amidasa permanece unida a los restos de colina del LTA. El LTA podría actuar *in vivo* como una barrera topológica, uniendo la enzima a través de sus residuos de colina e impidiendo el acceso a su sustrato como consecuencia de la formación de agregados micelares (Briese y Hakenbeck, 1985).

Al igual que con el LTA, la adición de altas concentraciones de colina (2%) o fosforilcolina a un cultivo de neumococo provoca la aparición de células con un fenotipo *Lyt⁻* (Briese y Hakenbeck, 1985). La inhibición por colina es un efecto de tipo no competitivo, lo que sugiere que la adsorción de la enzima a la pared celular tiene lugar a través de un sitio diferente al centro catalítico (Giudicelli y Tomasz, 1984). Esto último ha sido corroborado posteriormente, como veremos más adelante.

2.1.1.3- Regulación celular

Las autolisinas bacterianas son enzimas que deben ser translocadas a través de la membrana citoplásmica, para llegar a su lugar de acción, la pared celular. Por otro lado, estas enzimas, debido a su acción bacteriolítica, deben estar estrictamente reguladas. En el caso de *S. pneumoniae* esta regulación se lleva a cabo a nivel postraducciona.

Briese y Hakenbeck (1985) determinaron que en el citoplasma celular de *S. pneumoniae* la forma E de la amidasa *LytA* constituye un 40% de la amidasa celular, mientras que el resto del enzima se encontraba asociado a la membrana plasmática como la forma C. La enzima asociada podía liberarse con colina, lo que sugería una interacción específica con los residuos de colina que forman parte del LTA de la membrana. En células EA, la amidasa se localiza exclusivamente en la fracción soluble, aunque la cantidad detectada era similar a la presente en células que se multiplicaban en presencia de colina. La

forma E es el precursor de la forma C o proenzima (García y cols., 1985), mientras que la conversión a la forma C tendría lugar después del transporte a través de la membrana plasmática tras la interacción con los residuos de colina del LTA. Puesto que la actividad catalítica de la autolisina tiene un absoluto requerimiento por colina, la eliminación de la misma bajo la acción de la fosfocolinesterasa (Höltje y Tomasz, 1974) crearía zonas resistentes a la acción de la autolisina a la vez que afectaría al poder inhibidor del LTA.

La amidasa, en células de neumococo cultivadas en presencia de colina, se ha localizado mediante microscopía electrónica en el septo de separación de las células hijas, lo que sugiere la intervención de la autolisina en el proceso de división celular (Díaz y cols. 1989). Mediante análisis inmunoquímicos y estudios de fraccionamiento celular se observó que la amidasa producida en *E. coli* se localizaba en la periferia celular, bien soluble (15%) o bien asociada por interacciones débiles (65%) a la cara periplásmica de la membrana interna. Estos resultados permiten especular con la posibilidad de que la amidasa LytA en *S. pneumoniae* pudiera asociarse a la cara interna de la membrana celular, probablemente a través de la interacción entre la carga negativa de los fosfolípidos de membrana y alguno(s) de los 42 aminoácidos cargados positivamente de la enzima (García y cols., 1986b). Posteriormente, la proteína sería translocada a la cara externa de la membrana a través de un mecanismo que no implica el procesamiento de un péptido señal y que podría ser común para la secreción de otras autolisinas. La amidasa LytA, una vez localizada en la cara externa de la membrana celular interaccionaría con el LTA que contiene colina, lo que provocaría tanto la conversión de la enzima como la inhibición de su actividad por atrapamiento micelar (Briese y Hakenbeck, 1985). La amidasa se mantendría así regulada (inhibida) durante la mayor parte del crecimiento celular, y sólo localmente (en la región del septo de la célula) y en determinados momentos de dicho ciclo (durante la separación de las células hijas) la enzima quedaría libre de regulación y ejercería su actividad. La actividad amidásica se vería también libre del efecto inhibidor de LTA mediante alteraciones que ocurren en la célula al final de la fase estacionaria de crecimiento (ausencia de nutrientes) ó mediante la adición al medio de antibióticos (Tomasz, 1979; Horne y cols., 1977).

2.1.2- Características de la glucosaminidasa

El aislamiento del mutante M31, completamente delecionado en el gen *lytA*, permitió la detección de una nueva actividad lítica que fue caracterizada como una

glicosidasa (Sánchez-Puelles y cols. 1986a). La supresión de la fuerte actividad amidásica presente en todas las cepas de neumococo, hacia de la cepa M31 una cepa ideal para la purificación de la nueva actividad litica, paso esencial para la asignación de un papel biológico a esta nueva enzima. Dicha enzima litica fue purificada a homogeneidad electroforética asignándole un M_r de 64.000 (García y cols, 1989). La colina de los ácidos teicoicos de pared condiciona la actividad de la enzima al igual que se había descrito para la amidasa LytA (Giudicelli y Tomasz, 1984). Esta enzima es capaz de degradar las paredes que contienen colina pero no paredes con etanolamina y su actividad se inhibe en presencia de colina y de LTA, aunque no es reconocida por el antisero contra la LytA. Su temperatura óptima de actuación es 30°C aunque no se excluye la existencia de actividad a 37°C, temperatura óptima de crecimiento de neumococo. De hecho, la curación fenotípica de la cepa M31 con dicha enzima a 37° conduce a la lisis del cultivo al final de la fase estacionaria de crecimiento, revertiendo al fenotipo de la cepa salvaje. Estos resultados demuestran que la glicosidasa se comporta como una autolisina que se mantiene bajo regulación. Se ha sugerido que las autolisinas presentes en cantidades minoritarias intervienen en el remodelado de la pared celular durante el crecimiento, así como en la consecución del estado de competencia para la transformación genética (Kawamura y Shockman, 1983).

La clonación del gen que codifica para la glucosaminidasa así como la obtención de mutantes de la misma, permitiría elucidar el posible papel biológico desempeñado por esta enzima.

2.2- Características de la PspA (Proteína A de superficie de neumococo)

La proteína PspA se ha identificado como la única proteína de superficie de neumococo que posee propiedades inmunogénicas y de virulencia. Se encuentra presente en todas las cepas de neumococo analizadas presentando variaciones de tipo serológico y en relación a su peso molecular (Crain y cols., 1990; Waltman y cols., 1990). Esta proteína presenta un M_r deducido de la secuencia de nucleótidos de 65.380, el cual no coincide con el estimado por SDS-PAGE de 84.000 Da. La estructura deducida a partir de la secuencia de aminoácidos revela la existencia de tres dominios estructurales diferentes que confieren distintas propiedades a la proteína: un dominio cargado correspondiente a la parte N-

terminal de la proteína con una estructura secundaria predominantemente en α -hélice, un dominio rico en prolina y un dominio C-terminal constituido por diez repeticiones de 20 aminoácidos que termina en una cola hidrofóbica de 13 aminoácidos. Este último dominio presenta una alta homología con el dominio C-terminal de unión a colina de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos, estando estructurado de la misma manera, mientras que el dominio N-terminal presenta homología con las proteínas M del grupo A de *S. pyogenes*, proteínas de superficie que carecen de una región C-terminal con repeticiones en *tandem*. De ello se deduce que la PspA sería el resultado de la unión de dos dominios proteicos diferentes, procedente cada uno de ellos de una familia de proteínas distinta y que durante el curso de la evolución se habrían fusionado para constituir una nueva proteína funcional con un mecanismo de unión a la superficie bacteriana diferente al de otras proteínas de superficie, al carecer del típico dominio de anclaje que presentan estas últimas (Yother y Briles, 1992a). La expresión de proteínas PspA truncadas tanto en el dominio N-terminal como C-terminal ha permitido asignar propiedades a los diferentes dominios de la molécula. El análisis de las proteínas truncadas en la región repetida C-terminal demostró que la variabilidad antigénica va asociada al dominio N-terminal en α -hélice de la molécula, siendo ésta probablemente la parte expuesta en la superficie. Algunas de éstas proteínas fueron secretadas al medio de cultivo en vez de permanecer asociadas a la célula, requiriéndose la delección de al menos cinco de las repeticiones para evitar el anclaje de la PspA a la envoltura celular y ser secretada al exterior (Yother y Briles, 1992b). Respecto a la variabilidad en las masas moleculares de la proteína PspA en diferentes cepas, es atribuible tanto a la mitad C-terminal como N-terminal (Yother y cols., 1992b).

Se ha demostrado que PspA actúa como una proteína integral de membrana debido a la interacción de su C-terminal con los residuos de colina del LTA de la membrana, quedando el dominio de prolina dentro de la pared celular y el N-terminal expuesto en el exterior. Dicha interacción no requiere del extremo C-terminal hidrofóbico, pero sí es dependiente del número de repeticiones presente en la región C-terminal. Este tipo de interacción parece representar un nuevo mecanismo de anclaje de una proteína de superficie de un microorganismo Gram-positivo. Su mecanismo de unión es dependiente de colina al igual que ocurre con las enzimas líticas de *S. pneumoniae* y sus bacteriófagos y se ha visto que puede liberarse al exterior a partir de células, protoplastos ó membranas tras la adición de altas concentraciones de colina y no es capaz de anclarse en las envolturas de células

crecidas en un medio EA. La proteína no parece estar asociada con los ácidos teicoicos de pared, quizás porque los residuos de colina de los ácidos teicoicos no están tan accesibles a la molécula o bien porque la proteína PspA interacciona con el LTA u otras moléculas de la membrana. Puesto que la PspA posee un péptido señal, debería ser constitutivamente secretada al exterior, pero la interacción del dominio C-terminal con la colina del LTA hace que sea retenida en la superficie celular (Yother y White, 1994).

Este mecanismo inusual de unión a la superficie celular que exhibe la PspA puede ser el resultado de la evolución modular tal como ocurre con las enzimas líticas de *S. pneumoniae* y sus bacteriófagos.

2.3- Enzimas líticas de fagos que infectan *S. pneumoniae*

La primera enzima lítica fágica capaz de hidrolizar la pared celular de neumococo que fue aislada y purificada a homogeneidad electroforética fue la codificada por el fago Dp-1 y recibió el nombre de PAL (Phage-Associated Lysin) (García y cols., 1983a). Esta enzima tiene un peso molecular de 31.000 y se ha caracterizado como una *N*-acetilmuramil-L-alanil amidasa (García y cols., 1984). Para llevar a cabo su actividad catalítica requiere la presencia de residuos de colina en la pared y se inhibe en presencia de LTA. Sin embargo, a diferencia de LytA, su actividad se estimula por agentes reductores como el DTT o el β -mercaptoetanol (García y cols. 1983a,b,c), se inhibe parcialmente por la cardiolipina y no da reacción inmunológica cruzada con la forma E de la amidasa LytA, aunque sí con la forma C (García y cols., 1984). Por último, se ha demostrado que esta enzima interviene en la liberación de la descendencia fágica en mutantes de neumococo que carecen de amidasa LytA por lo que ambas enzimas estarían implicadas en dicho proceso (García y cols., 1984).

La segunda enzima lítica caracterizada fue la muramidasa del fago Cp-1. Fue purificada a homogeneidad electroforética presentando un M_r de 39.000. Esta enzima no necesita conversión para alcanzar su máxima actividad, pero requiere la presencia de colina en los ácidos teicoicos de la pared. La colina libre y el LTA inhiben de manera no competitiva la actividad de la enzima (García y cols., 1987). Las similitudes bioquímicas y funcionales entre la amidasa LytA del huésped y la lisozima codificada por el fago Cp-1, sirvieron de punto de referencia para proponer que los genes que codifican para estas enzimas presentarían homología a nivel de nucleótidos. Esta hipótesis fue analizada

experimentalmente a través de ensayos de hibridación (García y cols., 1988a) y posteriormente fue confirmada sirviéndose de estudios comparativos de las secuencias de aminoácidos de la amidasa de neumococo y la lisozima fágica. Se observó que los dominios C-terminales de estas enzimas eran prácticamente idénticos lo que llevó a sugerir que estas regiones eran responsables del reconocimiento de la pared celular. Para establecer las características moleculares de los genes líticos presentes en otros fagos que infectan neumococo, mediante técnicas de hibridación se llevó a cabo la clonación y secuenciación de los genes líticos *cpl9* del fago Cp-9 y *cpl7* del fago Cp-7, utilizando como sondas la secuencia nucleotídica del C-terminal de la LytA y la secuencia N-terminal de la muramidasa Cpl-1 respectivamente (García y cols., 1990). De la comparación a nivel de nucleótidos de los genes *cpl1*, *cpl7* y *cpl9* se pudo observar una mayor similitud entre los genes *cpl1* y *cpl9* que con el gen *cpl7*. Así pues, de los 600 primeros nucleótidos del gen *cpl9*, 52 eran diferentes (8,7%), existiendo tan sólo 20 cambios diferentes en los 417 nucleótidos restantes (4,8%). En el caso del gen *cpl7*, sólo el 17,3% de los 600 primeros nucleótidos es diferentes a los de *Cpl1*, mientras que con el resto del gen la divergencia es completa. Los genes *cpl9* y *cpl7*, codifican para lisozimas de 342 y 339 aminoácidos con masas moleculares de 38.461 y 39.107 Da respectivamente. Tan sólo 10 aminoácidos eran diferentes entre *Cpl1* y *Cpl9*, 9 de los cuales se localizaba en la región N-terminal y 1 sólo en la C-terminal. Los aminoácidos Asp-9 y Glu-36, que parecen estar implicados en los procesos catalíticos (Sanz y cols, 1992), se conservan tanto en *Cpl1* como en *Cpl7* y *Cpl9*. Mientras que en el caso del fago Cp-9, se trata de una lisozima dependiente de la presencia de colina para su actividad y posee dominios C-terminales homólogos a los de la LytA y *Cpl-1*, en el caso del fago Cp-7 se encontró un gen que codificaba para una lisozima capaz de degradar, *in vivo* e *in vitro*, paredes de neumococo donde la colina había sido reemplazada por etanolamina y cuyo dominio C-terminal era completamente diferente a los encontrados en las enzimas líticas anteriores. Estos resultados confirmaban la hipótesis de que los dominios C-terminales estaban implicados en el reconocimiento específico del sustrato.

También se ha clonado y secuenciado el gen lítico del fago atemperado HB-3 (Romero y cols., 1990a,b) utilizando el gen *lytA* como sonda. Es una enzima dependiente de colina y al igual que la LytA, requiere del proceso de conversión. Es uno de los casos de

mayor similitud a nivel de nucleótidos entre los genes de un fago y una bacteria (87%), la cual desaparece en las regiones no codificantes de dichos genes.

El estudio de estos genes ha proporcionado un sistema adecuado para investigar los requerimientos necesarios para la recombinación entre bacterias y DNAs de fagos. La combinación de las técnicas de transformación genética y de *Southern-blot*, han permitido demostrar que la integración del gen *hbl* en el cromosoma del huésped, depende del grado de similitud entre los DNAs del donador y del receptor.

Más recientemente se ha clonado y secuenciado el gen lítico del fago EJ-1 (Díaz y cols., 1992), otro fago atemperado aislado de una cepa clínica y atípica de neumococo que codifica también para una amidasa (*Ejl*) que depende de la presencia de colina en el sustrato que degrada y requiere del proceso de conversión. Posee un dominio C-terminal homólogo a los de la amidasa LytA, presentando un M_r de 36.535. Es interesante destacar que Ejl tiene una delección de 2 aminoácidos en el último módulo de colina de la región C-terminal equivalente al de las lisozimas Cpl1 y Cpl9. Presenta una similitud del 80% con LytA en la región estructural, existiendo una divergencia completa en las zonas flanqueantes.

2.3.1- Estructuración en dominios de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos

Conviene destacar que todos los dominios C-terminales están constituidos por seis motivos de secuencias repetidas, de unos 20 aminoácidos de longitud cada uno, a excepción del fago Cp-7, que posee 2,8 repeticiones exactamente iguales, de 48 aminoácidos cada una y diferentes a las anteriores (Fig. 4). Este hallazgo llevó a sugerir que la colina de la pared habría servido a lo largo del proceso evolutivo como un fuerte elemento de presión selectiva contribuyendo a la conservación de las regiones C-terminales de todas estas enzimas líticas, excepto Cpl7 (García y cols., 1988a).

Por otra parte, el análisis global de las enzimas líticas de neumococo y sus bacteriófagos permitió observar que las enzimas que poseían la misma actividad lítica contenían dominios N-terminales semejantes. Así, el dominio N-terminal de la amidasa LytA de neumococo era homólogo al de las amidasas Hbl y Ejl y, el dominio N-terminal de la lisozima Cpl1 presentaba homología con los dominios N-terminales de Cpl7 y Cpl9 (Fig. 4) (García y cols. 1990).

Teniendo en cuenta los datos anteriores se sugirió la existencia de una organización modular de las enzimas líticas del sistema de neumococo (Fig. 4) donde el dominio C-terminal estaría implicado en funciones de reconocimiento y unión al sustrato mientras que el centro activo se localizaría en el dominio N-terminal.

2.4- Enzimas líticas de pared de *C. acetobutylicum*

Puesto que la biomasa y la producción de solventes se ven afectados por las autolisinas como ya se destacó en el apartado 1.2.2, el estudio del sistema autolítico de *C. acetobutylicum* permitiría tener un mayor conocimiento de su papel fisiológico así como de las implicaciones económicas asociadas con el proceso de fermentación (Soucaille y Goma, 1986).

La enzima litica mejor caracterizada es la lisozima de *C. acetobutylicum* ATCC 824, la cual ha sido purificada, y el gen *lyc* que la codifica, clonado y expresado en *E. coli*. (Croux y cols., 1991, 1992a). La secuencia nucleotídica del gen *lyc* codifica para una proteína de 324 residuos con un M_r de 34.939. El análisis comparativo de la secuencia de aminoácidos, entre la muramidasa *Lyc* y otras enzimas líticas de pared, reveló la existencia de una importante similitud de su región N-terminal con las lisozimas de *Streptomyces globisporus*, el hongo *Chalaropsis* y la familia Cp de los bacteriófagos de *S. pneumoniae*. Los aminoácidos responsables de la actividad catalítica se localizan en dicho dominio y en posiciones equivalentes en todas estas proteínas. El análisis detallado de la secuencia de aminoácidos de la región C-terminal muestra la fusión de cinco posibles motivos repetidos, mientras que el análisis comparativo con las proteínas del banco de datos revela similitud con el dominio N-terminal de la muramoil-pentapéptido-carboxipeptidasa de *Streptomyces albus*, atribuyéndole a esta región un papel en el reconocimiento y unión al sustrato (Croux y García, 1991). Todos estos resultados se relacionaban bien con la ya conocida organización modular de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos (apartado 2.3.1), en donde la región N-terminal es la responsable de la actividad enzimática, mientras que el dominio C-terminal, estructurado en motivos repetidos, era la responsable del reconocimiento y unión al sustrato.

Webster y cols. (1981) describieron en *C. acetobutylicum* P262, una actividad litica de pared asociada a un glicoproteína de 28 kDa. Otra lisozima de 44 kDa se había

purificado de *C. acetobutylicum* NI-4, denominada anteriormente *C. sacharoperbutylaceticum* ATCC 27021 (Yoshino y cols., 1982). Una *N*-acetil-muramoyl-L-alanina amidasa de 100 kDa de peso molecular se aisló de un cultivo lisado de *C. acetobutylicum* NI-504 (*C. sacharoperbutylaceticum* ATCC 27022) infectado con el bacteriófago HM-7 (Hongo y cols., 1974; Ogata y cols., 1974). Posteriormente Podvin y cols. (1988) descubrieron en *C. acetobutylicum* NI-4, la existencia de una actividad lítica dependiente de colina así como la presencia de dicho aminoalcohol formando parte de los ácidos teicoicos de pared. La presencia de colina como componente de la pared bacteriana ya se había descrito en *S. pneumoniae*, como hemos visto en el apartado 1.1.2.1, demostrándose la participación de la misma en la regulación de las enzimas líticas del huésped y sus bacteriófagos, y actuando como elemento de presión selectiva contribuyendo así a la conservación de las regiones C-terminales de todas estas enzimas. Puesto que algunas especies de *Streptococcus* están filogenéticamente relacionadas con el género *Clostridium* (Ludwig y cols., 1985), en un intento de encontrar relaciones evolutivas entre las enzimas líticas de ambos microorganismos, García y cols. (1988b) fueron capaces de obtener de los sobrenadantes de las cepas *C. acetobutylicum* CECT806 (*C. acetobutylicum* NCIB8052) y *C. acetobutylicum* NI-504, una actividad lítica dependiente de colina. Dicha observación permitió purificar dicha actividad lítica en columnas de DEAE-celulosa. Las fracciones activas se analizaron a homogeneidad electroforética observándose la presencia de una banda con un M_r de 115.000 y cuya actividad correspondía a la de una *N*-acetil-muramyl-L-alanyl-amidasa. Aunque las estirpes NCIB 8052 y NI-504 se consideran especies distintas por mostrar diferentes características (Yoshino y cols., 1982), no se encuentran diferencias en sus actividades líticas aunque sí se presenta mayor cantidad de enzima lítica en la cepa de *C. sacharoperbutylaceticum*.

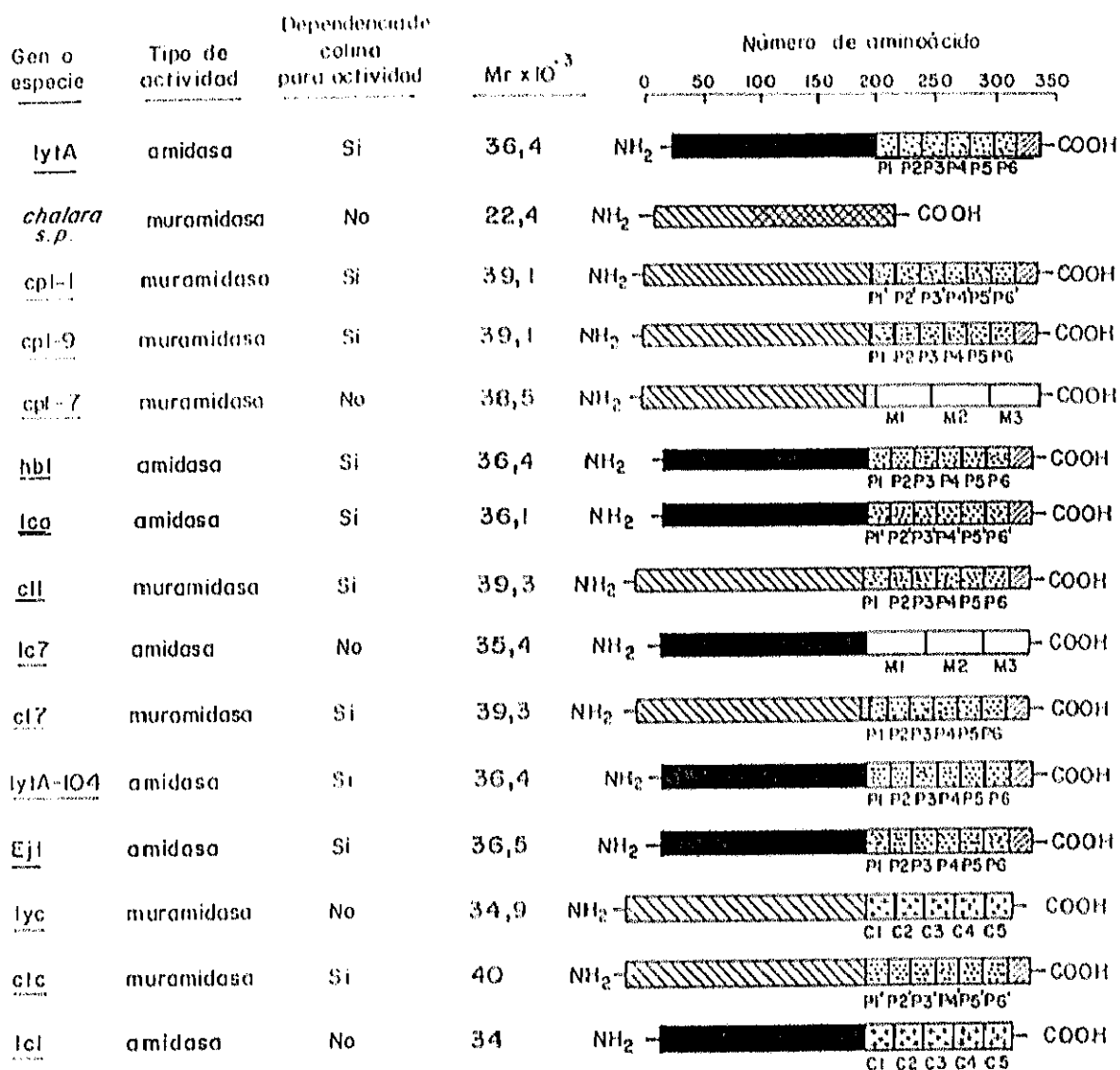


Figura 4- Representación esquemática de las enzimas líticas de neumococo y sus bacteriófagos, de *C. acetobutylicum* ATCC 824 y del hongo *Chalara*, así como de algunas quimeras construidas a partir de ellas. Las zonas de igual sombreado indican un alto grado de similitud. Los genes *lca* y *cll* son construcciones quiméricas desarrolladas a partir del gen *lytA* de neumococo y el gen *cpl1* del fago Cp-1. Los genes *lc7* y *cl7* son quimeras construidas entre los genes *lytA* de neumococo y *cpl7* del fago Cp-7. Los genes *clc* y *lcl* son quimeras construidas entre genes *lyc* de *C. acetobutylicum* ATCC 824 y *cpl1* del fago Cp-1. P1 a P6 representan las secuencias repetidas (motivos) de los dominios carboxilos de las enzimas *LytA*, *LytA101* o *Ejl*. P1' a P6', representan las secuencias repetidas (motivos) de los dominios carboxilos de las enzimas *Cpl1* o *Cpl9*. M1 a M3 y C1 a C5 representan las secuencias repetidas (motivos) de los dominios carboxilo de las enzimas *Cpl7* y *Lyc*, respectivamente.

3- TEORÍA MODULAR DE LA EVOLUCIÓN DE LAS PROTEÍNAS

Las teorías actuales sobre la evolución de las proteínas y los genes aceptan el concepto de "modularización", asumiendo que los nuevos genes y proteínas proceden no sólo de una sustitución puntual de un nucleótido o aminoácido sino del ensamblaje de secuencias de DNA preexistentes (Blake, 1978). Por lo tanto, las proteínas estarían estructuradas en unidades funcionales o módulos con una funcionalidad independiente del conjunto de la proteína de la cual forman parte. Este mecanismo evolutivo se basa en la captación ó intercambio de las secuencias génicas correspondientes a dichos módulos, de manera que un único suceso de recombinación entre dos de estos módulos podría originar un nuevo gen y por tanto una nueva proteína híbrida funcional (Gilbert, 1978). Todo esto permitiría una mejor adaptación de los organismos a las condiciones ambientales cambiantes. Esta teoría se puede aplicar también a los organismos eucarióticos debido a la especial distribución de sus genomas, permitiendo el reordenamiento de las secuencias de DNA y favoreciendo la aparición de proteínas híbridas construidas en base al intercambio de estos módulos.

Atendiendo al tamaño de la unidad fundamental o módulo, existen diferentes concepciones de esta teoría. Botstein (1980) propuso que los genomas fágicos se han generado a partir de módulos constituidos por agrupaciones de genes con una actividad específica. Mientras algunos autores mantienen la idea de que las proteínas proceden del intercambio de dominios, otros sugieren que los péptidos, codificados por cortas secuencias de DNA y con una conformación definida, constituyen los módulos en los que se encuentran estructuradas las proteínas.

Aunque se han realizado pocas aproximaciones experimentales para verificar esta teoría, sí se ha encontrado en *S. pneumoniae* un excelente modelo experimental para apoyar la hipótesis generalizada de que la mayor parte de las proteínas poseen una organización modular. Esto se basa en tres pruebas evidentes: i) el análisis comparativo de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos; ii) la expresión de proteínas truncadas tanto en el dominio C-terminal como en el N-terminal con funciones definidas; iii) la construcción de enzimas quiméricas funcionales fago-bacteria y entre bacterias de diferentes géneros.

Del primer aspecto ya se habló en el apartado 2.3.1, en donde se sugirió la organización modular de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos en base a

sus relaciones estructurales y funcionales. Según esto, surgió la idea de que los dominios estructurales de estas enzimas podrían ser intercambiables, propiciando de este modo la posibilidad de preparar enzimas quiméricas funcionales que combinaran las propiedades bioquímicas de las enzimas parentales. La demostración de la existencia de dos dominios independientes se realizó tras la clonación y expresión de las regiones de los genes *lytA* y *cpl-1*, que codifican para el dominio C-terminal (Sánchez-Puelles y cols., 1990), así como las regiones *lytA*, *cpl-1* y *cpl-7* que codifican el dominio N-terminal, de sus respectivas enzimas (Sanz y cols., 1992). Respecto al dominio C-terminal, se comprobó que mantenía su capacidad para el reconocimiento de colina en ausencia del N-terminal, pero no presentaba actividad lítica. El dominio N-terminal, sin embargo, era capaz de ejercer su función lítica, adoptando una conformación activa incluso en ausencia del correspondiente dominio de unión a colina. Además, era capaz de hidrolizar las paredes que contenían etanolamina incluso con una eficiencia mayor que las paredes con colina, lo que corrobora de nuevo la importancia del dominio de unión a colina en el reconocimiento del sustrato. Sin embargo, sus actividades específicas disminuían considerablemente en relación a las actividades de las enzimas completas.

Todos estos resultados se confirmaron posteriormente con la construcción de proteínas quiméricas activas, tanto entre dominios homólogos como no homólogos, favoreciendo aún más esta teoría. A partir de la comparación de la secuencia de aminoácidos se comprobó que existía una región intermedia entre los dominios N- y C-terminales que permitía introducir modificaciones por mutagénesis dirigida, sin alterar la actividad de la enzima así modificada (García y cols., 1990; Croux y García, 1991). Esto facilitó la construcción de genes quiméricos y demostró que se trataba de una región neutra desde el punto de vista funcional, como cabría esperar de una organización modular como la propuesta. En una primera aproximación se construyeron dos enzimas quiméricas entre los genes *lytA* y *cpl1* (Lca y CII) (Díaz y cols., 1990b) en base a la similitud que presentan las regiones C-terminales de las enzimas parentales (García y cols., 1988a) (Fig. 4). Sin embargo, las quimeras construidas entre la amidasa *LytA* y la lisozima *Cpl7* (Lc7 y C17), se originaron mediante el intercambio de dominios no homólogos (Fig. 4). En ambos casos se obtuvieron proteínas quiméricas activas, cuyas propiedades bioquímicas, además de ser una combinación de las que presentaban las enzimas parentales, podían atribuirse a los dominios concretos que habían sido intercambiados (Fig. 4). Estas enzimas quiméricas eran activas *in*

vivo cuando se ensayaban en el sistema homólogo mediante experimentos de curación fenotípica e incluso mediante la expresión de los genes en la cepa de neumococo M31, utilizando plásmidos capaces de replicarse en este sistema (Díaz y cols., 1991).

Por otro lado, la construcción de enzimas líticas quiméricas activas entre diferentes géneros bacterianos confiere un mayor apoyo a la teoría modular postulada anteriormente, pues los procesos de recombinación genética no tienen por qué tener lugar sólo entre un fago y su huésped, sino incluso entre microorganismos heterólogos. La observación de que algunas estirpes bacterianas de *C. acetobutylicum* poseían colina como componente de la pared (apartado 1.2) condujo al estudio a nivel molecular, de la lisozima de *C. acetobutylicum* ATCC 824 que, aunque carecía de colina, presentaba una organización modular equivalente a la descrita para las enzimas líticas de neumococo (Croux y García, 1991). De esta manera, se llevó a cabo la construcción de dos enzimas quiméricas, una primera entre el dominio C-terminal de la lisozima del fago Cp-1 y el dominio N-terminal de la enzima lítica de *C. acetobutylicum* (Clc) (Croux y cols., 1993a) y otra, entre el dominio N-terminal de la amidasa LytA y la región C-terminal de la lisozima de *C. acetobutylicum* (Lcl) (Croux y cols., 1993b). Mientras que en el primer caso las enzimas parentales presentan dominios homólogos, en el segundo, es la primera vez que se describe la construcción de una quimera lítica activa a partir de la fusión de genes carentes de cualquier homología y procedentes de distintos géneros bacterianos.

Los estudios realizados con todas estas proteínas quiméricas activas, demostraron que el centro activo de las enzimas líticas de pared tanto de *S. pneumoniae* y sus bacteriófagos como de *C. acetobutylicum*, se localiza en el dominio N-terminal de la molécula. Las quimeras Lca, Lc7 y Lcl, eran amidasas (Fig. 4) porque su dominio N-terminal correspondía al de la amidasa LytA. Recíprocamente, las quimeras Cll, Cl7 y Clc eran lisozimas puesto que su dominio N-terminal correspondía al de las lisozimas parentales Cpl1, Cpl7 y Lyc respectivamente.

Por otra parte, estas construcciones también demuestran que el dominio C-terminal es el responsable del reconocimiento del sustrato. Así, las quimeras Cll, Lca, Cl7 y Clc, las cuales presentan el dominio de unión a colina de las enzimas líticas parentales, presentan una actividad dependiente de la presencia de colina en los ácidos teicoicos de pared, a la vez que se inhiben en presencia de dicho aminoalcohol. Además, las enzimas quiméricas Cl7 y Cll cuyo dominio de unión a colina procede de la amidasa LytA, requieren el proceso de

conversión (Fig. 4). Sin embargo, la quimera Lc7 degrada indistintamente paredes con colina o con etanolamina y su actividad no se inhibe específicamente en presencia de colina, como cabría esperar al proceder su dominio C-terminal de la lisozima Cpl7. La quimera Lcl, es un caso excepcional, pues no fue capaz de degradar las paredes celulares de *C. acetobutylicum*, a pesar de poseer el dominio C-terminal que reconoce las mismas. Esto puede ser debido a que la composición y/o conformación de su peptidoglicano no permite la correcta orientación o estabilización entre las unidades de mureína y el centro activo de la enzima. El pH óptimo sobre paredes que contienen colina, ya sean de *S. pneumoniae* o de *C. acetobutylicum* NCIB 8052, era idéntico al de la enzima parental Lye, implicando por tanto al dominio C-terminal en las propiedades catalíticas atribuidas al dominio N-terminal.

Puesto que los dominios N- y C-terminales conservan su funcionalidad independientemente uno del otro y a la vez son intercambiables, cabe considerarlos como módulos, según los criterios propuestos por Doolittle (1985). La baja actividad específica que presentan los dominios N-terminales aislados hace suponer que dichos módulos pudieron ser en un tiempo enzimas funcionales, en las cuales, la adquisición del dominio de reconocimiento de colina les confirió una ventaja evolutiva al mejorar el sistema de reconocimiento del sustrato insoluble, consiguiendo así una mayor eficiencia catalítica (es el caso de la quimera Cl7). Por otro lado, los resultados obtenidos con las enzimas quiméricas permiten concluir que el dominio C-terminal de unión a colina no es responsable sólo del reconocimiento del sustrato sino del proceso de conversión y de la inhibición de la actividad enzimática en presencia de colina. Además, ejerce cierta influencia sobre las propiedades catalíticas atribuidas al dominio N-terminal en función del pH.

4- OTRAS PROTEÍNAS CON MOTIVOS SIMILARES A LOS DE UNIÓN A COLINA

Análisis recientes de datos de secuencias de aminoácidos han permitido identificar otras clases de proteínas bacterianas, las cuales, al igual que las enzimas líticas de *S. pneumoniae* y sus bacteriófagos, se caracterizan por su capacidad de unirse a una gran variedad de ligandos, mostrando secuencias conservadas repetidas localizadas en la región C-terminal de la molécula. Dichas repeticiones se pueden encontrar en las toxinas de *C. difficile* (Tox A y Tox B) y en las glucosiltransferasas y proteínas de unión a glucanos de *Streptococcus* (Tabla 3). A pesar de las diferencias funcionales entre estas proteínas, la similitud detectada parece estar relacionada con el reconocimiento y unión a receptores glicosídicos.

Tabla 3- Proteínas con dominios de secuencia conocida, homólogos al de unión a colina, de *Streptococcus* y *Clostridium*.

Cepa Bacteriana	Enzima	Clase	Referencia
<i>C. difficile</i> VPI 10463	Tox A		Dove y cols. (1990)
	Tox B		Von Eichel-Streiber y Sauerborn (1990)
			Barroso y cols. (1990)
BARTS W1	Tox A		Von Eichel-Streiber y cols. (1992a)
			Wren y cols. (1990)
<i>S. mutans</i> GS-5	GTF-B	GTF-I	Shiroza y cols. (1987)
	GTF-C	GTF-SI	Ueda y cols. (1988)
	GTF-D	GTF-S	Honda y cols. (1990)
Ingbritt	GRP		Hanas y cols. (1990)
<i>S. downei</i> Mfe 28	GTF-I	GTF-I	Ferreti y cols. (1987)
	GTF-S	GTF-S	Gilmore y cols. (1990)
<i>S. sobrinus</i> OM2176	GTF-T	GTF-S	Hanada y cols. (1993)
<i>S. salivarius</i> ATCC 25975	GTF-J		Giffard y cols. (1991)
	GTF-K	GTF-S	Giffard y cols. (1993)

4.1- Estructura primaria de las toxinas de *C. difficile*.

C. difficile, el microorganismo responsable de la colitis pseudomembranosa y la colitis asociada a antibióticos en humanos (Bartlett y cols., 1980), produce dos toxinas

relacionadas: la toxina A, una enterotoxina y la toxina B, una potente citotoxina, considerada como el mayor determinante de virulencia de este microorganismo. Ambas toxinas son secretadas, habiendo sido aisladas del medio de cultivo (Taylor y cols., 1981).

Se han clonado y secuenciado los genes de las toxinas A y B de algunas cepas de *C. difficile* (Dove y cols., 1990; von Eichel-Streiber cols., 1990; Wren y cols., 1990; Barroso y cols., 1990). Estas toxinas poseen un M_r deducido de la secuencia de aminoácidos de 308 y 269 kDa respectivamente. Ambas proteínas presentan una similitud de un 49% con un 14% de sustituciones conservadas (von Eichel-Streiber y cols., 1992a). Dichas proteínas inducen anticuerpos policlonales específicos mostrando diferente reacción tanto *in vivo* como *in vitro*. En estas moléculas se pueden distinguir dos dominios. El dominio N-terminal posee tres regiones claramente diferenciadas: (i) una zona hidrofóbica de 172 aminoácidos, altamente conservada en ambas proteínas y cuya función se desconoce; (ii) una zona rica en residuos no presentes en otras partes de la molécula (arginina, cisteína, histidina y triptófano), la cual parece desempeñar un papel biológico importante; (iii) una región de baja homología entre ambas proteínas y un dominio C-terminal con motivos repetidos. El dominio C-terminal de la toxina A de la cepa VPI 10463 contiene 38 unidades repetidas y continuas, que ocupan 853 residuos de un total de 2710 (Dove y cols., 1990). Dichos motivos repetidos se han clasificado, según muestra la tabla 4, en dos clases principales: siete repeticiones de clase I compuestas de 30 aminoácidos, con una similitud de un 71-100% entre cada dos repeticiones. Las 31 repeticiones de clase II muestran una similitud más limitada con las de clase I y tienen 20-21 aminoácidos de longitud (Wren y cols., 1991). Las repeticiones de clase II se han subdividido en cuatro subclases: IIA (7 copias), IIB (13 copias), IIC (5 copias), y un grupo más variado en cuanto a homología IID (6 copias). Todas las repeticiones contienen el dipéptido tirosina-fenilalanina que se ha utilizado para alinear las secuencias consenso de la toxina A, con los motivos repetidos de las otras proteínas de *Clostridium* y *Streptococcus*. La secuencia C-terminal de la toxina A de una cepa diferente de *C. difficile* (BARTS W1) contiene un número de repeticiones similares a los de la cepa anterior, pero con algunas diferencias (Wren y cols., 1990).

El dominio repetido de la toxina A es altamente inmunogénico y se ha visto, mediante estudios de hemoaglutinación, que es responsable de la unión a carbohidratos (Price y cols., 1987; Lysterly y cols., 1990). Krivan y cols. (1987) encontraron que el ligando receptor de la toxina A contiene un trisacárido terminal de Gal α 1-3Gal β 1-4GlcNAc.

Recientemente se ha definido la porción de hemoaglutinación de la toxina A utilizando un péptido sintético (TIDGKKYYFN) situado en la repetición de una unidad de la clase IIB.

Tabla 4- Caracterización de los motivos repetidos consenso de los dominios de unión a ligando de proteínas de los géneros *Streptococcus* y *Clostridium*.

CEPA BACTERIANA	PROTEINA	TAMAÑO AA	MOTIVOS REPETIDOS			
			CLASE	Nº AA	Nº Rep.	Secuencia consenso de aa
<i>C. difficile</i> VPI 10463	Tox A	2,710	I	30	7	HQIQVEKGPNGFEYFAPANYDANNIEGQAI VYQNKFLTNGKKYYFGNNS KAVTGWQTI DGKKYYFNPTA IASTQYKTI DGKHFFDT DGI IAAxGxxTIXNxxYYFxxDxx HQIQVENTEDGFKYFAPANTLDENxEGEAI xxxxGxxxIDxxKYYFxxxGA
			IIA	20	7	
			IIB	21	13	
			IIC	21	5	
			IID	21	6	
	Tox B	2,366	BI	30	4	
			BIIE	20	20	
<i>S. mutans</i> GS-5	GTF-B	1,475	A	33	6	WYFDDNNOYAVTOARTINGQHLYFRANGVQVKG HGRISYYIDGNSGDQIRNRFV WFLFDHNOXAVTOAVxxxNGxxxYFxxPNGxQAKN DGKLRVFDxxSGNQxxxNRV WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
	GTF-C	1,375	C	20	4	
	GTF-D	1,430	A	33	4	
			C	20	4	
			A	33	5,5	
Ingbritt	GBP	563	C	20	4	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	34	5,5	
<i>S. downei</i> Mfe 28	GTF-I	1,597	C	20	2	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	35	6	
			B	48	2	
	GTF-B	1,365	C	20	5	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	32	6	
<i>S. sobrinus</i> OM2176	GTF-T	1,468	C	20	4	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	34	4,5	
<i>S. salivarius</i> ATCC 25975	GTF-J		C	21	5	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	33	5,5	
	GTF-K		C	20	4	WYFDSNOVAVTQxQTI DGxxxYFxxSNGKQVKG NGKxxxYFxxAxSGELAVNRFA
			A	35	7	
			D	12	5	
<i>S. pneumoniae</i> Fagos Cp-1 Cp-9	LytA			21	6	GWxKxxDxWYFDDxxGxMxx xWxKxxGxWYFDDxxGxMxx xWxKxxGxWYFDDxxGxMxx
	Cp11			21	6	
	Cp19			21	6	

X= cualquier aminoácido

AA= aminoácidos.

Rep= Repeticiones.

Barroso y cols. (1990), han publicado la secuencia de nucleótidos del gen de la toxina B. El análisis en su composición de aminoácidos revela que parte de la región C-terminal de la toxina contiene 24 motivos repetidos, los cuales están claramente relacionados con la toxina A en tamaño, secuencia y alineamiento. Se han subdividido en clase I y clase II al igual que en la toxina A (Tabla 4) (Wren y cols., 1991). De clase I existen cuatro motivos repetidos de 30 aminoácidos, denominados BI. Estos motivos están muy relacionados con los de clase I de la toxina A. Por otro lado, presenta 20 motivos repetidos de clase II (BIIE), de 20 ó 21 aminoácidos, los cuales no se encuentran relacionados con los de la misma subclase en la toxina A. Aunque todavía no se conoce el receptor de esta toxina, es bastante probable que estas secuencias repetidas estén implicadas también en el reconocimiento del sustrato. Es posible que las toxinas presenten diferentes receptores puesto que la toxina A presenta una actividad hemoaglutinante mientras que la toxina B, no.

4.2- Estructura primaria de las glicosiltransferasas y proteínas de unión a glucanos de *Streptococcus*

Los patógenos orales *S. mutants*, *S. downei* y *S. salivarius*, tienen la habilidad de sintetizar proteínas extracelulares, incluyendo las proteínas de unión a glucanos (GBPs) y las glicosiltransferasas (GTFs). Estas proteínas se caracterizan por ser altamente hidrofílicas y presentar un péptido señal típico de proteínas secretables. Parecen ser responsables de la virulencia de los estreptococos orales, estando implicadas en el metabolismo de azúcares y en la subsecuente formación de placas y caries dentales. Se desconoce aún el mecanismo por el cual dichas proteínas interaccionan con la superficie celular del microorganismo para llevar a cabo la formación de las placas dentales.

La GBP de *S. mutants* es una proteína de 563 residuos con un M_r de 59.000 tras ser procesada. Contiene dos series de repeticiones no adyacentes designadas A y C, las cuales se extienden a lo largo de la proteína (Tabla 4) (Banas y cols., 1990). Dentro de las de tipo A, existen 5 repeticiones de 32-34 aminoácidos y una parcial de 12 aminoácidos, mientras que de las de tipo C existen 4 repeticiones de 17 a 20 aminoácidos. Los motivos de clase A presentan una gran similitud con los motivos repetidos de clase IIB de la toxina A, siendo ésta superior a la que se observa con cualquiera de las otras clases de repeticiones de las toxinas A y B e incluso presentan reacciones inmunológicas cruzadas. Esta GBP, carece de

actividad enzimática, constituyendo exclusivamente un dominio de unión a carbohidratos. Parece estar implicada en la formación de placas dentales puesto que un mutante defectivo de la misma pierde la capacidad de adherirse *in vitro*.

Las GTFs son enzimas que tienen un M_r promedio de 160.000 y son responsables de la hidrólisis de los restos de sacarosa, utilizando los restos de glucosa liberados para la polimerización de cadenas glicánicas tanto solubles como insolubles en agua, que van a formar parte de la placa dental (Hanada y cols., 1980). Las GTFs se clasifican en tres grupos: (i) Las GTFs que producen polímeros solubles en agua, donde las unidades se unen mediante enlaces α -1,6 (GTF-S); (ii) GTFs que producen polímeros insolubles en agua, con uniones α -1,3 (GTF-I); (iii) GTFs que sintetizan una mezcla de ambos tipos de polímeros (GTF-SI). Se sabe que las cadenas glicánicas insolubles desempeñan un papel importante en la formación de la placa dental debido a su facilidad para adherirse a superficies rugosas mediando la agregación de las células bacterianas y los restos de comida. El papel de los polímeros solubles se desconoce aún, aunque parece que intervienen en los procesos de agregación y pueden servir como una fuente de energía extracelular. Se ha comprobado que muchas GTF-Is requieren de un cebador para comenzar con la síntesis de la cadena glucídica, mientras que ciertas GTF-Ss son independientes de cebador (Yamashita y cols., 1989) y por tanto podrían actuar como iniciadores para la síntesis mediada por las GTF-Is.

Las GTFs presentan también repeticiones en el dominio C-terminal de la molécula. Las mejor caracterizadas son las de clase A con una media de 33 aminoácidos por repetición y las de clase C, similares a la GBP de *S. mutants*, con una media de 20 aminoácidos. Hasta el momento se han secuenciado los genes que codifican para 8 GTFs (Tablas 3 y 4). Hay que destacar que la GTF-I de *S. downei* presenta además dos repeticiones de tipo B y la GTF-K de *S. salivarius*, 5 de tipo D exclusivas de estas proteínas. Los genes *gtfB* y *gtfC* de *S. mutants* GS-5 se encuentran próximos en el cromosoma y parecen constituir un operón. Los genes que codifican para la GTF-J y la GTF-K de *S. salivarius* también se encuentran próximos en el cromosoma, pero a pesar de la gran similitud de las GTFs de ambos microorganismos, presentan diferentes actividades siendo la GTF-J y la GTF-K enzimas independientes de cebador mientras que la GTF-B y GTF-C son dependientes. El porcentaje de identidad entre las repeticiones de clase A y clase C de las GTFs y la GBP oscila entre un 38-58% y un 40-60%, respectivamente. Por otro lado, Ferreti y cols (1987) han comprobado que deleciones en la región C-terminal de

las GTF-I de *S. downei* eliminan la capacidad de unión de la enzima a su sustrato, lo que sugiere que esta zona es el dominio de unión a carbohidratos.

Las repeticiones A y C muestran poca similitud entre ellas y las posiciones relativas de las repeticiones C varían entre unas proteínas y otras. La comparación de secuencia de aminoácidos entre todas las repeticiones del tipo A de estas proteínas, las toxinas de *C. difficile* y los motivos de unión a colina de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos, revela la existencia de 3 aminoácidos conservados en todos los casos, el dipéptido tirosina-fenilalanina y una glicina (Tabla 4). En el caso de las repeticiones C, conservan también en algunos casos el dipéptido tirosina-fenilalanina y en su defecto el dipéptido tirosina-tirosina.

5- OBJETIVOS

La colina es un aminoalcohol que forma parte de los lípidos de membrana de las células eucarióticas, siendo poco frecuente su presencia en procariotas (Arondel y cols., 1993). En la introducción de esta Tesis se han descrito diferentes proteínas producidas por bacterias y fagos que poseen capacidad de unión a colina. Sin embargo, a pesar de la presencia de colina en diferentes microorganismos, sólo se han clonado algunos genes que codifican para estas proteínas, de *S. pneumoniae* y sus bacteriófagos, permitiéndonos elucidar la estructura del dominio de unión a colina y extraer algunas conclusiones interesantes sobre el papel evolutivo que desempeña este aminoalcohol. Dichos genes codifican para enzimas líticas, a excepción del antígeno PspA cuya función en neumococo se desconoce por el momento.

El estudio comparativo de las secuencias de aminoácidos de las enzimas líticas de neumococo y sus bacteriófagos, nos ha permitido obtener información relativa a su peculiar organización modular. Estas enzimas son el resultado de la fusión de dos dominios independientes: el dominio N-terminal que contiene el centro catalítico y proporciona la especificidad enzimática, y el dominio C-terminal, responsable de la unión de la proteína a componentes específicos de la pared celular. La construcción de proteínas quiméricas, mediante intercambio de dominios entre estas enzimas líticas, demostró que la adquisición del dominio de unión a colina representa una ventaja evolutiva para aquellas enzimas que interaccionan con un sustrato polimérico, permitiendo conseguir una mayor eficiencia catalítica. Se ha sugerido que la presencia de colina actúa como un elemento de presión selectiva al evitar, o al menos limitar, la divergencia evolutiva natural del dominio de unión a dicho ligando, determinando la creación de quimeras funcionales. Esto implica que los microorganismos que poseen colina como componente de la pared, o de la membrana en forma de LTA, deberían contener proteínas con dominios similares, lo que vendría a corroborar dicha hipótesis. La caracterización de otras proteínas con dominios de unión a colina y, funciones diferentes a las descritas, apoyarían aún más la teoría modular.

La presencia de colina en la pared de algunas estirpes de *C. acetobutylicum* junto con la detección de una actividad lítica de pared, dependiente de dicho aminoalcohol, en los sobrenadantes de cultivo de algunas de estas estirpes (García y cols., 1988b), nos llevó a la

búsqueda de relaciones evolutivas entre los dominios de unión a colina de proteínas presentes en *C. acetobutylicum* y en *S. pneumoniae* y, al estudio de su estructura modular.

Por todo ello, los objetivos de esta Tesis Doctoral son los siguientes:

- I) Caracterización molecular de genes que codifican para proteínas con dominios de unión a colina en *C. acetobutylicum* NCIB 8052.
- II) Caracterización molecular de genes que codifican para nuevas proteínas con dominios de unión a colina en *S. pneumoniae*.

II- MATERIALES Y MÉTODOS

1- ESTIRPES BACTERIANAS, BACTERIOFAGOS Y PLÁSMIDOS

Para el desarrollo de este trabajo se utilizaron las estirpes bacterianas, los bacteriófagos y los plásmidos que se detallan en la Tabla 5.

Tabla 5- Estirpes bacterianas, bacteriófagos y plásmidos

Estirpe, fago o plásmido	Genotipo o fenotipo relevante	Referencia/Fuente
<i>E. coli</i>		
TG-1	F ⁺ (<i>lacI^r lacZΔM15</i>)	Gibson, 1984
DH5α	(Φ80 <i>lacZΔM15</i>) <i>endA1</i> <i>recA1</i> , <i>hsdR17</i> (<i>r_k⁻ m_k⁻</i>) <i>supE44 ΔlacU169</i>	Hanahan, 1983
C600	<i>thi-1</i> , <i>leuB</i> , <i>thr-1</i>	Appleyard, 1954
<i>C. acetobutylicum</i>		
<i>C. acetobutylicum</i> ATCC 824		Croux y cols., 1992b
<i>C. acetobutylicum</i> NCIB 8052		García y cols., 1988
<i>C. acetobutylicum</i> NI-4 ^a		Hongo, 1960
<i>C. acetobutylicum</i> NI-504		Hongo y cols, 1965
<i>S. pneumoniae</i>		
R6	estirpe salvaje	Rockefeller Univ.
M24	Lyt ⁻ S3 ⁻	García y cols., 1993
M31	Hex ⁻ Δ <i>lytA</i>	Sánchez-Puelles, 1986a
406 ^b	Lyt ⁺ S3 ⁺	J. Casal
13868 ^b	Lyt ⁺ S1 ⁺	A. Fenoll
12928 ^b	Lyt ⁺ S6 ⁺	Arrecubieta y cols., 1994
13783 ^b	Lyt ⁺ S14 ⁺	Arrecubieta y cols., 1994
13805 ^b	Lyt ⁺ S9 ⁺	Arrecubieta y cols., 1994
8249 ^b	Lyt ⁺ S19A ⁺	Arrecubieta y cols., 1994
12950 ^b	Lyt ⁺ S23 ⁺	Arrecubieta y cols., 1994

Tabla 5 Continuación

Estirpes, fagos o plásmidos	Genotipo o fenotipo relevante	Referencia/Fuente
<i>S. pneumoniae</i>		
D39	Lyt ⁺ S2	Bracco y cols., 1957
Fagos		
Cp-1		Ronda y cols., 1981
Plásmidos		
pUC18	Ap ^R	Vieira y Messing, 1982
pUC19	Ap ^R	Yanisch-Perron y cols., 1985
pINIII-(lpp ^P -5)-A3	Ap ^R	Inouye e Inouye, 1985
pUCE191	MLS ^R , Tc ^R	E. García
pLSE3	MLS ^R , Tc ^R	Díaz y García, 1990a
pLysS	Cm ^R	Studier, 1991

^a Esta cepa fue denominada *C. sacharoperbutylacetonicum* anteriormente

^b aislados clínicos del serogrupo o serotipo indicado

2- MEDIOS Y CONDICIONES DE CULTIVO

Las cepas se se conservaron congeladas a -70 °C en medio de cultivo al que se añadió glicerol al 10% (v/v) para el caso de *E. coli* y *S. pneumoniae* y del 15% (v/v) para el caso de *C. acetobutylicum*. En el momento de sembrarlas se descongelaron, incubándose en los medios correspondientes a 37°C, a menos que se indique lo contrario, con agitación para las cepas de *E. coli* y sin agitación para las cepas de *S. pneumoniae*. Las cepas de

Clostridium se crecieron a esa misma temperatura, sin agitación y en condiciones anaerobias en campana de anaerobiosis modelo Forma Scientific.

2.1- *E. coli*

Para el cultivo de *E. coli* se utilizó el medio LB descrito por Sambrook y cols. (1989), añadiendo agar al 1,5 % (p/v) para cultivos en medio sólido. La concentración de antibiótico añadida al medio LB para el cultivo de cepas resistentes fue de 100 µg/ml para la ampicilina, 10 µg/ml para la tetraciclina, y 20-30 µg/ml para el cloranfenicol. El crecimiento de las cepas de *E. coli* se siguió por turbidimetría a 550 nm con un espectrofotómetro Shimadzu UV-260.

2.2- *S. pneumoniae*

Para el cultivo de *S. pneumoniae* se utilizó el medio C + Y: medio CpH8 (Tomasz y Hotchkiss, 1964) suplementado con extracto de levadura 0,08% (p/v) (Difco). Para cultivos en medio sólido se añadió al medio anterior, agar al 1% (p/v) y catalasa (250 unidades/ml) (Boehringer Mannheim) aunque en otros caso se utilizaron placas de agar sangre [Tryptic-Soy agar (Difco), suplementado con 5% de sangre desfibrinada de carnero]

Las cepas transformantes de neumococo que contienen derivados de pLSE1 (Ronda y cols., 1988) se seleccionaron en un medio con tetraciclina a 0,8 µg/ml (medio líquido) ó 1 µg/ml (medio sólido) ó bien eritromicina (10 µg/ml) ó lincomicina (1 µg/ml).

El crecimiento de este microorganismo se siguió por nefelometría con el empleo de un nefelómetro Coleman modelo 9, calibrado para contaje de viables en neumococo donde $1 \text{ N} = 7,5 \times 10^5$ unidades formadoras de colonias (UFC/ml).

2.3- *C. acetobutylicum*

Para el cultivo de este microorganismo se utilizaron los medios CGM (Roos y cols., 1985), TYA (Ogata y Hongo, 1973), 2XYTG (Oultram y cols, 1988) y RCM (Reinforced Clostridial medium, Difco). Las esporas se activaron por calor a 70-80 °C durante 15 minutos. El crecimiento se siguió por turbidimetría a 600 nm con un espectrofotómetro Shimadzu UV-260.

3- PROCEDIMIENTOS DE TRANSFORMACIÓN GENÉTICA

3.1- *E. coli*

Este microorganismo se transformó utilizando el método del RbCl (Kushner, 1978) o bien mediante electroporación (Dower y cols., 1988).

3.2- *S. pneumoniae*

La preparación de cepas competentes de neumococo ha sido descrito por Tomasz (1970). Los cultivos de neumococo poseen competencia natural en una determinada etapa del crecimiento (la comprendida entre 280 N y 300 N. Una vez que el cultivo alcanza esta concentración celular, se le añade glicerol al 10% y se congela rápidamente. Las células conservadas a -70 °C mantienen la competencia sin pérdida apreciable durante largos periodos de tiempo.

El método de transformación utilizado fue esencialmente el descrito por Barany y Tomasz (1980), excepto que la incubación con el DNA transformante se realizó a 30°C durante 40 minutos en medio CpH8 suplementado con seroalbúmina bovina al 0,08% (p/v). Después los cultivos se incubaron a 37°C durante 120 minutos en medio C + Y con seroalbúmina bovina al 0,08% (fase de expresión), antes de la siembra a 37°C en el medio sólido adecuado.

4- ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA-SDS

Las electroforesis analíticas de proteínas se realizaron en todos los casos en condiciones desnaturalizantes en presencia de dodecilsulfato sódico (SDS-PAGE). La técnica descrita fue la utilizada por Laemmli (1970) utilizando geles de poliacrilamida en placa (160 x 100 x 2) entre el 8 y el 15 %. Las muestras se hirvieron durante 5 minutos en presencia del tampón de ruptura (Tris-HCl 62,5 mM, pH 6,8; SDS al 2%, β-mercaptoetanol al 5%, glicerol al 10% y azul de bromofenol al 0,005%). Las electroforesis se realizaron a temperatura ambiente y corriente constante (50 mA), utilizando un electrolito que contenía Tris-HCl 0,025 M, glicina 0,192 M y SDS al 0,1%. Las proteínas de los geles se visualizaron con azul brillante de Coomassie R-250, según se describe en Swank y Munkress (1971). Las proteínas empleadas como marcadores de tamaño molecular se

adquirieron de Bio-Rad: fosforilasa b (92,5 kDa); seroalbúmina bovina (66,2 kDa); ovoalbúmina (43 kDa); anhidrasa carbónica (31 kDa); inhibidor de tripsina (21,5 kDa).

5- OBTENCIÓN DE EXTRACTOS CRUDOS DE CULTIVOS CELULARES

Para obtener extractos de *E. coli* y *S. pneumoniae*, se cultivaron estos microorganismos en 10 ml del medio indicado en cada caso a 37°C hasta alcanzar la fase exponencial media. Los cultivos se enfriaron a 4°C y se centrifugaron a 6000 x g durante 10 minutos, resuspendiéndose las células en 5 ml de tampón fosfato sódico 20 mM, pH 6,9. Una vez homogeneizada la suspensión, las células se rompieron en un sonicador MSE modelo MK2 mediante cuatro tratamientos de 10 segundos cada uno, manteniéndose siempre la muestra a 4°C. La suspensión sonicada se centrifugó a 4°C y 10.000 x g durante 10 minutos para eliminar los restos celulares. El sobrenadante de esta centrifugación (extracto crudo) se conservó a -20°C. La fracción proteica se valoró mediante el método de Bradford (1976) y se analizó mediante SDS-PAGE.

6- PURIFICACIÓN DE LA PROTEÍNA CspA

6.1- *C. acetobutylicum*

La proteína CspA se purificó a partir del sobrenadante de cultivo de la cepa de *C. acetobutylicum* NCIB 8052 mediante cromatografía de afinidad en columnas de DEAE-celulosa tal como describen García y cols. (1988b). Puesto que la proteína CspA no se presentaba pura, aunque constituía el 90% del extracto purificado, se procedió a la purificación de la misma mediante electroelución utilizando un electroluidor "Bio-Rad 422 Electroeluter Model". Para ello, se llevó a cabo un SDS-PAGE en un gel al 8% con el extracto purificado en DEAE-celulosa. La tinción del gel se realizó con el método de tinción reversa con sales de imidazol-zinc (Fernández-Patrón y cols., 1992). La banda de proteína mayoritaria correspondiente a la proteína CspA se recortó y se electroeluyó siguiendo las instrucciones de los proveedores.

6.2- *E. coli*

La proteína CspA se purificó a partir de un extracto celular de un cultivo de *E. coli* TG-1 (pKDI547 y pLysS) en fase exponencial de crecimiento. El pLysS es un plásmido que tiene clonado el gen de la lisozima del fago T7 que infecta *E. coli*, el cual, tras la congelación y descongelación de las células provoca la lisis de las mismas, sin necesidad de recurrir a ningún procedimiento mecánico de ruptura. En nuestro caso, las células se resuspendieron y concentraron veinte veces en tampón fosfato sódico 20 mM pH 6,9 con NaCl 0,5 M, previa congelación y descongelación de las mismas para provocar su lisis. El extracto se sonicó con la finalidad de romper el DNA y se centrifugó a 90.000 x g durante treinta minutos con el fin de eliminar las envueltas celulares y precipitar el DNA tras la adición de estreptomicina al 2%. El sobrenadante se depositó en una columna de DEAE-celulosa equilibrada previamente con tampón fosfato sódico 20 mM pH 6,9, realizándose el proceso a 4°C. Después de lavar la columna con 40 volúmenes del mismo tampón fosfato anterior más NaCl 1,5 M, la proteína retenida por interacción específica con la colina se eluyó con el mismo tampón pero con NaCl 1,5 M y colina 140 mM. La fracción proteica se detectó mediante el método de Bradford (1976) y se analizó en un SDS-PAGE.

7- ENSAYOS DE ACTIVIDAD ENZIMÁTICA

Se realizó sobre paredes de celulares de *S. pneumoniae* o de *C. acetobutylicum* NCIB 8052, según el caso, marcadas radiactivamente con [*metil*-³H] colina (60 Ci/mmol; Amersham)].

7.1- Ensayo enzimático de los extractos obtenidos de *E. coli*

Se utilizó el método clásico para la determinación de la actividad amidásica de neumococo (Höltje y Tomasz, 1976), con algunas modificaciones. La mezcla de reacción contiene 200 µl del tampón fosfato sódico 20 mM pH 6,9 a los que se añaden 10 µl de paredes celulares de neumococo marcadas radiactivamente y de 5 a 30 µl de la dilución adecuada del extracto que se desea ensayar. La incubación se realiza a 37°C a diferentes tiempos y la reacción se detiene tras la adición de 10 µl de formaldehído al 35% y 10 µl de seroalbúmina bovina al 4%. Las muestras se centrifugan en una microfuga Hettich a 10.000 x g durante 10 minutos. A 150 µl del sobrenadante de cada muestra se les añade 3,5 ml de

líquido de centelleo (Cocktail-22, Normascint, Scharlau) y se determina la radiactividad en un contador de centelleo líquido LKB Wallac (modelo 1219 Racbeta). Una unidad de actividad enzimática se define como la cantidad de enzima que cataliza la solubilización de 1 μg de paredes en 10 minutos (Höltje y Tomasz, 1976).

7.2- Ensayo enzimático de los extractos obtenidos de *S. pneumoniae*

La mezcla de reacción que contiene 200 μl del tampón fosfato sódico 20 mM pH 6,9 con Brij al 0,1% y de 5 a 30 μl de la dilución adecuada del extracto que se desea ensayar se mantiene 5 minutos a temperatura ambiente para que el detergente solubilice las membranas. Después se añaden 10 μl de paredes celulares de neumococo marcadas radiactivamente y se incuba a 37°C ó 30°C a diferentes tiempos. El resto es igual que en el caso anterior.

7.3- Ensayo enzimático de las proteínas purificadas

La mezcla de reacción contiene 200 μl de tampón fosfato sódico 20 mM pH 6,9, 10 μl de paredes y 10 μl de la dilución adecuada de la enzima purificada y se incuba a diferentes tiempos. El resto es igual al primer caso.

8- TÉCNICA DEL WESTERN BLOT

Las proteínas purificadas y las proteínas de los extractos crudos, una vez separadas mediante electroforesis en geles de poliacrilamida-SDS, se transfirieron a membranas de nitrocelulosa (Schleicher and Schuell) según describen Sambrook y cols. (1989). La membrana se saturó con leche en polvo desnatada al 5% (p/v) en tampón PBS (fosfato sódico 10 mM, pH 7,4; NaCl 140 mM) mediante incubación a 4°C durante 12 horas. Posteriormente, se incubó a temperatura ambiente con agitación suave durante 4 horas en presencia del suero anti-CspA. Después de tres lavados de 10 minutos de duración cada uno con PBS que contenía Tween al 0,1% (v/v), la membrana se hizo reaccionar durante 1 hora a temperatura ambiente y suave agitación con anticuerpos de cabra anti-IgGs de conejo conjugados con peroxidasa (Jackson Immunoresearch). Finalmente, tras un lavado con tampón PBS, las bandas de reacción de los anticuerpos se visualizaron con peróxido de hidrógeno y 4-cloro-1-naftol (Sigma).

9- OBTENCION DE ANTICUERPOS ANTI-CspA

La obtención de anticuerpos policlonales dirigidos contra la proteína CspA se llevó a cabo mediante la inmunización de conejos con la proteína CspA purificada mediante electroelución tal y como se describe en el apartado 5.1. Se utilizaron dos dosis de 0,28 µg de proteína en solución salina. La primera en adyuvante de Freund completo (1:1; v/v) y, la segunda, en adyuvante de Freund incompleto (1:1; v/v), dejando transcurrir un mes entre ambas. La administración del antígeno se realizó de forma subcutánea. Una semana después de la administración de la segunda dosis, se extrajo sangre a los conejos para la obtención del suero. La sangre se incubó a 37°C durante 5 horas y después se mantuvo a 4 °C durante 12 horas para permitir la formación del coágulo. El sobrenadante se centrifugó a 10.000 x g 5 minutos a 4°C, para eliminar restos celulares y, por último, se inactivaron las proteínas del complemento durante 15 minutos a 56°C. Se ensayó la especificidad del suero obtenido, así como la posible reacción inespecífica del suero pre-inmune, mediante análisis por *Western-blot* utilizando diferentes diluciones del mismo. Para evitar la detección de bandas inespecíficas en el *Western-blot*, formadas por la posible reacción cruzada del suero con otras proteínas celulares diferentes de la CspA, se preincubó el suero a 37°C durante al menos una hora con un extracto celular de la cepa que posteriormente se iba a emplear para expresar la proteína CspA, pero con el plásmido control sin el gen que codifica para dicha proteína.

10- FRACCIONAMIENTO SUBCELULAR EN *C. acetobutylicum* NCIB 8052

De un cultivo de *C. acetobutylicum* NCIB 8052 en fase exponencial se centrifugaron 10 ml a 4°C, lavándose posteriormente las células dos veces con 20 mM de tampón fosfato sódico pH 6,9 y conservándose el sobrenadante del medio de cultivo a -20°C. El precipitado de células se resuspendió en 7 ml del mismo tampón y se rompieron en un sonicador MSE modelo MK2 mediante cuatro pulsos de 15 segundos cada uno, manteniendo siempre la muestra a 4°C. La suspensión sonicada conteniendo la fracción soluble (citoplasma) y la sedimentable (envolturas celulares), se centrifugó durante 1 hora a 100.000 x g. El sedimento con las envolturas celulares se lavó dos veces con el tampón fosfato mencionado anteriormente, resuspendiéndose posteriormente en 2 ml del mismo

tampón. La concentración de proteína de cada fracción se determinó mediante el método del Bradford (1976), utilizando la curva patrón de la seroalbúmina bovina.

11- ANÁLISIS DE CARBOHIDRATOS

Los extractos purificados en columnas de DEAE-celulosa de los sobrenadantes de *C. acetobutylicum* NCIB 8052 se corrieron en un gel SDS-PAGE al 10% y se tiñeron con un kit de detección de glicoproteínas suministrado por Boehringer Mannheim (DIG Glycan Detection Kit) tras la inmovilización de las proteínas en membranas de nitrocelulosa. Para la determinación cualitativa y cuantitativa de carbohidratos de la proteína CspA, la proteína electroeluida se hidrolizó con ácido sulfúrico 4 N a 100°C durante 5 horas, y se neutralizó con carbonato bórico. Los azúcares neutros, convertidos en los correspondientes acetatos de alditol, se identificaron y cuantificaron por GLPC (Guarro y cols., 1993) utilizándose inositol como patrón interno.

12- DETERMINACIÓN DE LA SECUENCIA N-TERMINAL DE LA PROTEÍNA CspA.

La secuencia de aminoácidos del extremo N-terminal de las proteínas CspA y CspC se determinó mediante el método de degradación de Edman, según se describe en López y cols. (1982a). Para ello se utilizó un secuenciador automático modelo 477A (Applied Biosystems).

13- EXTRACCIÓN DE DNA CROMOSÓMICO

13.1- *C. acetobutylicum*

Los DNAs de cepas de *Clostridium* que figuran en la tabla 5 se obtuvieron de la siguiente manera: las células procedentes de un cultivo de 400 ml en fase de crecimiento exponencial ($DO_{550}=1-1.3$), se centrifugaron y se lavaron con tampón TEG (50 mM glucosa, 25 mM Tris-HCl, pH 8; 10 mM EDTA). Posteriormente, las células se resuspendieron en una solución de lisis (sacarosa 10%; SDS 1%), manteniéndose a 37°C durante 20 minutos. A continuación se añadió proteinasa K a una concentración de 50

$\mu\text{g/ml}$ y EDTA 10 mM, incubándose durante 2 horas más a 55-60°C. Después de fenolizar varias veces dicha solución, el DNA cromosómico se precipitó con 0,6 volúmenes de isopropanol, se lavó con etanol al 70%, se secó y se resuspendió en aproximadamente 1 ml de tampón TE (10 mM Tris-HCl, pH 8; 1 mM EDTA).

13.2- *S. pneumoniae*

Para la extracción de DNA cromosómico de las cepas de neumococo Lyt⁻ se partió de 200 ml de un cultivo exponencial (600-700 N) crecido en medio C+Y, se centrifugaron y se lavaron con tampón fosfato 20 mM pH 6,9, resuspendiéndose posteriormente en 8 ml del mismo tampón. A continuación se curaron (Tomasz y Waks, 1975) tras añadir 10 ml de una solución de amidasa pura a una concentración de 2 mg/ml, incubándose 10 minutos a 37 °C para producir la lisis celular. Posteriormente se adicionó 0,5 ml de EDTA 0,5 M y se mantuvo 10 minutos más a 37°C. Tras la adición de otros 0,5 ml de EDTA 0,5 M y 1 ml de DOC al 10% e incubación durante 15 minutos más a 37 °C, se añadieron 0,25 ml de SDS al 20% y proteinasa K a una concentración final de 100 $\mu\text{g/ml}$. Dicha solución se incubó durante 1 hora a 37°C, ultracentrifugándose el lisado resultante en un gradiente de ClCs/BrEt (Sambrook y cols., 1989).

La concentración de DNA se estimó, en ambos casos, a partir de la absorbancia de la preparación a 260 nm. Una unidad de absorbancia equivale a una concentración de DNA de 50 $\mu\text{g/ml}$.

14- AISLAMIENTO DE RNA

14.1- Aislamiento de RNA de *C. acetobutylicum* NCIB 8052

Para el aislamiento de RNA total se utilizó el procedimiento descrito por Walter y cols. (1992) con alguna modificación. La cepa de *C. acetobutylicum* NCIB 8052 se incubó en 300 ml del medio TYA tras la adición de un preinóculo multiplicado durante la noche en el mismo medio. Se fueron tomando muestras de 10 ml a diferentes tiempos a lo largo de la curva de crecimiento. Se provocó la lisis celular con tres ciclos sucesivos de congelación y descongelación con nieve carbónica y posterior mantenimiento a -70 °C durante 20 minutos. Los lisados se resuspendieron en 3 ml del tampón AE (20 mM de acetato sódico pH 5,3 y 1

mM EDTA). A dicha suspensión se añadieron 6 ml de fenol ácido y SDS al 0,3 % previamente calentado a 60°C y se mantuvo agitando durante 6 minutos a la misma temperatura. Las muestras se mantuvieron 5 minutos a 4°C y se centrifugaron en frío, recogiendo la fase superior, la cual se ajustó a 0,25 M con acetato sódico 2 M pH 5,3. Se llevaron a cabo tres extracciones más con fenol ácido y se precipitó el RNA con 2,5 volúmenes de etanol absoluto, se secó y se disolvió en TE 10:10 (10 mM Tris-HCl, pH 8; 10 mM EDTA).

14.2- Aislamiento de RNA de *S. pneumoniae*

La cepa R6 de *S. pneumoniae* se incubó en medio C + Y y se fueron tomando muestras del cultivo a lo largo de la curva de crecimiento. Las muestras se centrifugaron y lavaron con TE (10:10). Las células se lisaron con TE conteniendo 0,1% de DOC, congelándose y descongelándose dos veces sucesivas con nieve carbónica. Posteriormente, las células se incubaron a 60 °C durante 5 minutos con una solución que contenía SDS al 0,5% (p/v), acetato sódico pH 4.0 y fenol ácido equilibrado con agua. Las muestras se centrifugaron recuperándose la fase acuosa y extrayéndose el RNA con fenol caliente y posterior precipitación con etanol absoluto. El RNA extraído se disolvió en el volumen adecuado de TE (10:10).

En ambos casos el DNA presente en las muestras se eliminó mediante una hora de tratamiento a 37°C con DNAsa I libre de RNAsa (Pharmacia) en presencia de un inhibidor de RNAsas (RNasina de Boehringer). La concentración de RNA se estimó a partir de la medida de absorbancia a 260 nm, sabiendo que una unidad de absorbancia a 260 nm equivalen a una concentración de 40 µg/ml. Para determinar la calidad del RNA, las muestras se analizaron en un gel de agarosa al 1%. Las muestras se conservaron a -70.

15- ELECTROFORESIS DE DNA

15.1- Electroforesis convencional

Se utilizaron geles de agarosa al 0,7 ó 1% en tampón TAE (Tris-HCl 40 mM, ácido acético 20 mM, EDTA 2 mM, pH 8,1), utilizando el mismo tampón como electrolito. A las muestras se les añadió 1-1,5/10 de su volumen de una solución compuesta por Ficoll 400 al 30% (p/v), azul de bromofenol al 0,2% (p/v), xilencianol al 0,2% y EDTA 40 mM pH 8,0.

La electroforesis se realizó a 100-150 V durante 60-90 minutos, y una vez finalizada los geles se tiñeron con bromuro de etidio (5 µg/ml) para su visualización con radiación ultravioleta. Como marcador de tamaño, se utilizó el DNA del fago lambda digerido con la enzima de restricción *BstEII* (Amersham).

15.2- Electroforesis en campo pulsado (PFGE)

La preparación del DNA, embebido en los bloques de agarosa, se llevó a cabo según Gasc y cols. (1991) para *S. pneumoniae* y según Wilkinson y Young (1993) para *C. acetobutylicum*. El DNA se digirió con las enzimas de restricción adecuadas según el caso, previamente a su electroforesis en tampón TBE (Tris 75 mM, ácido bórico 25 mM y EDTA 0,1 mM). La rampa utilizada fue de 1 a 50 s durante 24 horas para el caso de *S. pneumoniae* y de 5 a 50 s durante 20 horas para el caso de *C. acetobutylicum*, con un voltaje en ambos casos de 200 V. Tras la electroforesis, el gel se tiñó con una solución de bromuro de etidio (1 µg/ml) durante 20 minutos, haciéndose posteriormente una foto del gel. Los fragmentos de DNA se transfirieron a membranas de nylon siguiendo la técnica del *Southern-blot*, hibridándose posteriormente con las sondas radiactivas adecuadas a cada caso.

16- PREPARACIÓN DE PLÁSMIDOS

La preparación de plásmidos a partir de las cepas de *E. coli* se realizó utilizando el método de Birnboim y Doly (1979) tal y como lo describen Sambrook y cols. (1989).

17- MANIPULACIÓN DEL DNA CON ENZIMAS DE USO COMÚN EN BIOLOGÍA MOLECULAR

Las endonucleasas de restricción se obtuvieron de Boehringer Mannheim, Amersham, Pharmacia y New England Biolabs. La fosfatasa alcalina de intestino de ternera se obtuvo de Boehringer Mannheim. Amersham suministró la DNA ligasa y la polinucleótido quinasa del fago T4, así como el fragmento Klenow de la DNA polimerasa I de *E. coli*. Las enzimas y sus tampones correspondientes se utilizaron siguiendo las indicaciones recomendadas por las casas suministradoras.

18- AISLAMIENTO Y PURIFICACIÓN DE FRAGMENTOS DE DNA.

Los fragmentos resultantes de la digestión del DNA con las enzimas de restricción o los obtenidos mediante PCR, se aislaron y purificaron por tres procedimientos distintos.

18.1- Geles de poliacrilamida

Se utilizaron geles de poliacrilamida no desnaturalizantes preparados en TBE (Tris-borato 89 mM, ácido bórico 89 mM y EDTA 2 mM, pH 8,0) a una concentración variable dependiendo del tamaño de fragmento a aislar. Se empleó como electrolito para la electroforesis el mismo tampón TBE. A las muestras se les añadió 1/5-1/10 de su volumen de una solución compuesta por Ficoll 400 al 30% (p/v), azul de bromofenol al 0,2% (p/v), xilencianol al 0,2% y EDTA 40 mM pH 8,0. Una vez terminada la electroforesis, los geles se tiñeron con azul de metileno y se recortó la banda de DNA deseada en cada caso. El DNA se eluyó de la banda de acrilamida tal como describen Sambrook y cols. (1989).

18.2- Técnica del GeneClean

Los fragmentos que se desean aislar se separan primero mediante una electroforesis en geles de agarosa. Posteriormente, se recortan las bandas de interés y el DNA se purifica de la agarosa mediante su adsorción a diminutas partículas de vidrio siguiendo la técnica de GeneClean, tal como recomiendan los fabricantes (Bi 101 Inc., La Jolla, CA, USA).

18.3- Técnica de la β -Agarasa

El empleo de esta técnica también requiere la separación previa de los fragmentos mediante electroforesis en geles de agarosa de bajo punto de fusión y posterior tratamiento de la banda extraída con la β -Agarasa, que hidroliza la agarosa, liberando el DNA. Para ello es necesario fundir previamente la banda de agarosa recortada en el tampón adecuado (10 mM Bis tris-HCl pH 6,5; 10 mM EDTA) antes del tratamiento con la β -agarasa durante una hora a 40-42°C. Tras tratar con fenol y precipitar el DNA, éste se disuelve en el volumen adecuado de TE. Tanto la enzima como el tampón se obtuvieron de BioLabs (New England, USA).

19- SECUENCIACIÓN DE DNA

La secuenciación de DNA se realizó según el método de Sanger y cols. (1977). Para la reacción de secuenciación se utilizó el T7 SequencingTM Kit (Pharmacia) y como nucleótido marcado radiactivamente, [α -³⁵S]dATP (Amersham), siguiendo las recomendaciones de los fabricantes.

20- TÉCNICAS DE HIBRIDACIÓN DEL DNA y RNA

Los fragmentos de DNA marcados radiactivamente (sondas) se obtuvieron mediante la técnica del random-primer utilizando [α -³²P]dCTP (400 Ci/mmol, 10 μ Ci/ μ l) (Amersham), el fragmento Klenow, la solución de iniciadores (Pharmacia) y los dNTPs según se indica en Sambrook y cols. (1989). Todas las hibridaciones y lavados se hicieron entre 65 y 37°C según las circunstancias.

El marcaje de oligonucleótidos sintéticos se llevo a cabo mediante la utilización de la T4 polinucleótido quinasa utilizando como isótopo [γ -³²P] ATP (5000 Ci/mmol, 10 μ Ci/ μ l) siguiendo las indicaciones de Sambrooks y cols. (1989). Las hibridaciones se realizaron entre 37-40 °C y los lavados se hicieron sucesivamente a mayor temperatura, dependiendo del oligonucleótido utilizado como sonda y de la radiactividad incorporada en las membranas.

20.1- Southern-blot

La hibridación del DNA mediante la técnica descrita por Southern (1975), se realizó siguiendo el protocolo de Sambrook y cols. (1989). Las membranas de nylon para la transferencia del DNA se obtuvieron de Schleicher and Schuell. Las bandas radiactivas se detectaron utilizando películas HyperfilmTM-MP (Amersham) y pantallas amplificadoras Dupont, Cronex Lightning Plus, a -70°C. La deshibridación de las membranas se realizó con un primer tratamiento con NaOH 0,4 M durante 30 minutos a 45°C, y un segundo tratamiento con 0,1 x SSC (NaCl 0,3 M, citrato sódico 0,003 M) y SDS 0,1% otros 30 minutos a la misma temperatura.

20.2- Hibridación de DNA sobre filtros de colonias celulares

Las colonias de *E. coli* se transfirieron a filtros de nitrocelulosa (Millipore HATF, tamaño de poro 0,45 μm) como se indica en Sambrook y cols. (1989). Las células se lisaron sobre el filtro mediante tratamientos sucesivos, de 5 minutos de duración cada uno, con las soluciones que se indican: i) NaCl 1,5 M y NaOH 0,5 M (2 veces); ii) Tris-HCl 1M, pH 7,0; iii) Tris-HCl 0,5 M, pH 7,0 y NaCl 1,5 M; iv) 2 x SSC (NaCl 0,3 M, citrato sódico 0,003 M). La fijación del DNA al filtro se realizó mediante incubación a 80 °C durante 2 horas. Para eliminar los restos celulares, los filtros se hirvieron en agua destilada durante 10 minutos. La hibridación con la sonda radiactiva se realizó como se describe en Sambrook y cols. (1989).

20.3- Northern-blot

Tras el aislamiento de RNA, tal como se indica en el apartado 12, las muestras se cargaron en geles de agarosa con formaldehído y se corrieron a 9 V/cm durante la noche. El gel se dializó varias veces contra agua y se trató según describen Sambrook y cols. (1989). La transferencia de RNA se realizó mediante vacío utilizando como tampón 20 x SSC. La membrana se secó durante 30 minutos a 65°C y, posteriormente, se fijó el RNA con radiación ultravioleta 10 minutos.

20- TÉCNICA DE PRIMER EXTENSION PARA LA DETERMINACIÓN DE LOS SITIOS DE INICIACIÓN DE LA TRANSCRIPCIÓN

El RNA (19 μg) y el oligonucleótido marcado radiactivamente en su extremo 5' con $\gamma^{32}\text{P}$ ATP (10^6 cpm) se precipitaron conjuntamente, resuspendiéndose el pellet en 30 μl de buffer de hibridación (40 mM PIPES pH 6,4, 1 mM EDTA pH 8; 0,4 M NaCl y 50% formamida), desnaturalizando con calor a 85°C durante 10 minutos e incubando 12 horas a 30°C. El híbrido RNA-oligonucleótido se precipitó con etanol y se disolvió en 10 μl de la mezcla de reacción de la transcriptasa inversa [1x tampón AMV (Promega); 0,9 mM dNTPs pH 7; 0,4 U de RNAsina, y 3 U de enzima (avian myeloblastosis virus, Promega)]. La reacción de extensión transcurre durante una hora a 42°C y se detiene con NaOH a una concentración final de 0,4 M, manteniéndose 12 horas a temperatura ambiente para desnaturalizar los ácidos nucleicos. Posteriormente se neutralizó la mezcla con ácido acético

0,4 M y se precipitó con dos volúmenes de etanol absoluto, disolviéndose en una solución 1:1 (v/v) de agua y solución de colorante (azul de bromofenol al 0,3%, xilencianol al 0,3%, 10 mM de EDTA pH 7,5 y 97,5 % de formamida desionizada). Para determinar el sitio de iniciación de la transcripción se llevo a cabo una electroforesis en un gel de poliacrilamida al 6% conteniendo urea 8 M utilizando como referencia una reacción de secuenciación del plásmido que contiene la zona promotora de interés. El resultado se determina por autorradiografía.

22- REACCIÓN EN CADENA DE LA POLIMERASA

Para llevar a cabo la amplificación se empleó un equipo de Pharmacia LKB y las enzimas *Pfu* DNA polimerasa (Stratagene) y Replitherm™ DNA polimerasa termoestable (Epicentre). Las condiciones de amplificación se especifican en cada caso. Los productos amplificados se purificaron por la técnica de la β -Agarasa.

III- RESULTADOS

Clostridium acetobutylicum

1- CARACTERIZACIÓN MOLECULAR DEL GEN ESTRUCTURAL *cspA*

La proteína CspA (*Clostridium* *Secretable* *Protein*) se caracteriza por ser la proteína más abundante en el medio de cultivo de *C. acetobutylicum* NCIB 8052. Esta proteína ha sido purificada mediante cromatografía de afinidad en columnas de DEAE-celulosa, conteniendo el extracto purificado más del 90% de la proteína CspA (García y cols., 1988). Dicha preparación presentaba la peculiaridad de mostrar una actividad lítica de pared dependiente de colina, la cual se caracterizó como una amidasa (García y cols., 1988). Por ello parecía interesante la caracterización bioquímica y molecular de dicha proteína ya que podría tratarse de una autolisina dependiente de colina.

1.1- Clonación del gen *cspA*

Puesto que en el extracto purificado en DEAE-celulosa, la proteína CspA no estaba totalmente pura, se procedió a su purificación mediante electroelución siguiendo el procedimiento que se indica en el apartado 6.1 de Materiales y Métodos (Fig. 10). A partir de la proteína pura se llevó a cabo la determinación de la secuencia de aminoácidos de su extremo amino terminal (Fig. 5). Esta secuencia se utilizó para la síntesis de los oligonucleótidos degenerados, denominados O1 y O2 (Fig. 5), que se utilizaron como sondas para el análisis de varias genotecas construidas a partir del DNA cromosómico de *C. acetobutylicum* NCIB 8052 digerido con diversas enzimas de restricción y ligado al vector de clonación pUC18, utilizando *E. coli* TG-1 como huésped. La selección de las enzimas de restricción más adecuadas para construir genotecas se realizó en base a los resultados de experimentos de *Southern-blot*, utilizando como sondas oligonucleótidos sintéticos u otros fragmentos clonados. De este primer análisis se obtuvo un clon positivo que contenía el plásmido recombinante pKIT10 (Fig. 5) que portaba un fragmento *Pst*I de 279 pb y que codificaba para la región N-terminal de la proteína CspA. La utilización de este fragmento *Pst*I como sonda permitió la obtención de un nuevo plásmido recombinante denominado pKDI543 (Fig. 11) que contenía tres fragmentos *Dra*I aislados a partir de una genoteca *Dra*I. Mediante técnicas de hibridación se determinó el fragmento *Dra*I que contenía el gen *cspA*, el cual se aisló, purificó y se subclonó en pUC18 para dar lugar al plásmido pKDE543 (Fig. 5).

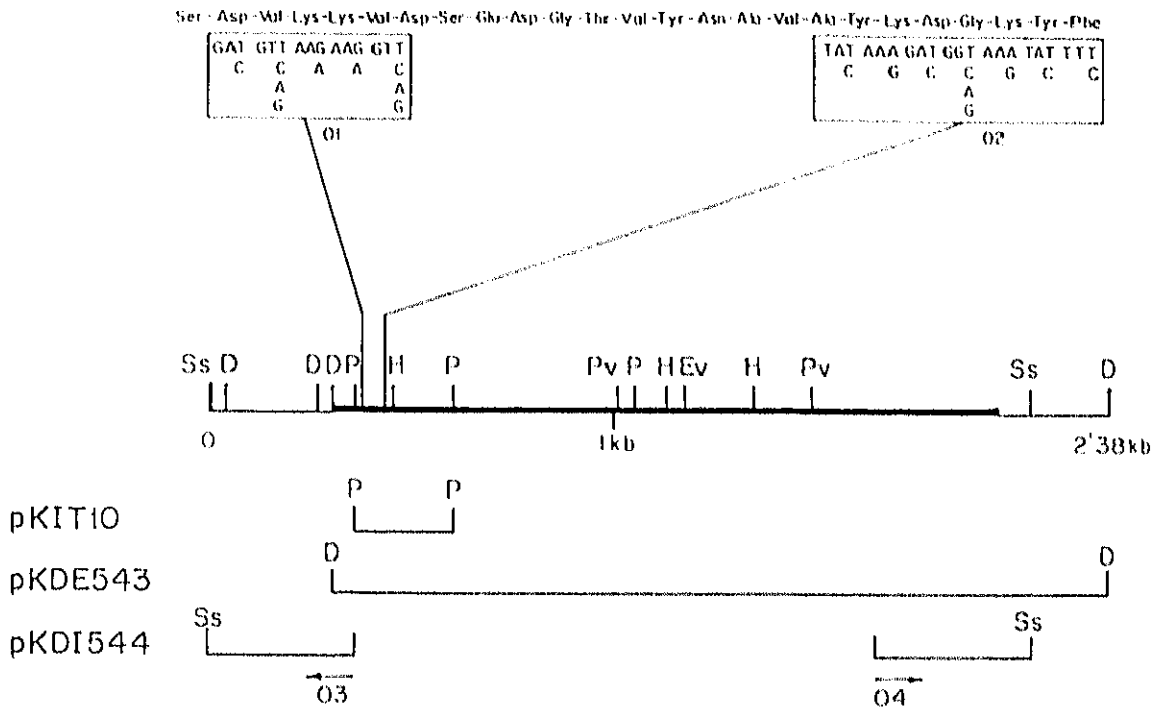


Figura 5- Mapa físico del gen *cspA*. En la parte superior de la figura se muestra la secuencia de aminoácidos de la región N-terminal de la proteína CspA en código de tres letras. O1 y O2 son los oligonucleótidos sintéticos utilizados en el primer análisis de genotecas. El trazo grueso negro corresponde al gen estructural *cspA* y el trazo fino a las regiones 5' y 3' no codificantes. Los insertos de los plásmidos pKIT10, pKDE543 y pKDI544, derivados de pUC18, se ilustran en la parte inferior de la figura. O3 y O4 son los oligonucleótidos sintéticos utilizados para la amplificación por PCR de la región 5' del gen *cspA*. Dianas de restricción: D, *DraI*; H, *HindIII*; Ev, *EcoRV*; Pv, *PvuII*; P, *PstI*; Ss, *SspI*. La longitud de los fragmentos se indica en kb.

Puesto que el gen aún estaba incompleto, como se pudo comprobar por su secuencia, la clonación del extremo 5' del gen se llevó a cabo mediante la técnica del PCR inverso (IPCR) (Hanada y cols., 1993), siguiendo la estrategia que se muestra en la figura 6. Puesto que conocíamos que el fragmento *DraI* ya clonado contenía sólo un sitio *SspI* situado después del bucle de terminación y que por *Southern-blot* el tamaño del fragmento *SspI* contenía el gen completo, se sintetizaron dos oligonucleótidos para amplificar mediante

PCR el resto del fragmento *SspI* que nos faltaba, utilizando como molde fragmentos de DNA digeridos con *SspI* y religados consigo mismo (Fig. 6). De esta manera se amplificó un fragmento de aproximadamente 700 pb que se ligó a pUC18 digerido con *SmaI*, transformándose posteriormente *E. coli* TG-1. Consecuentemente se obtuvo el plásmido pKDI544 que contiene la región 5' del gen *espA*.

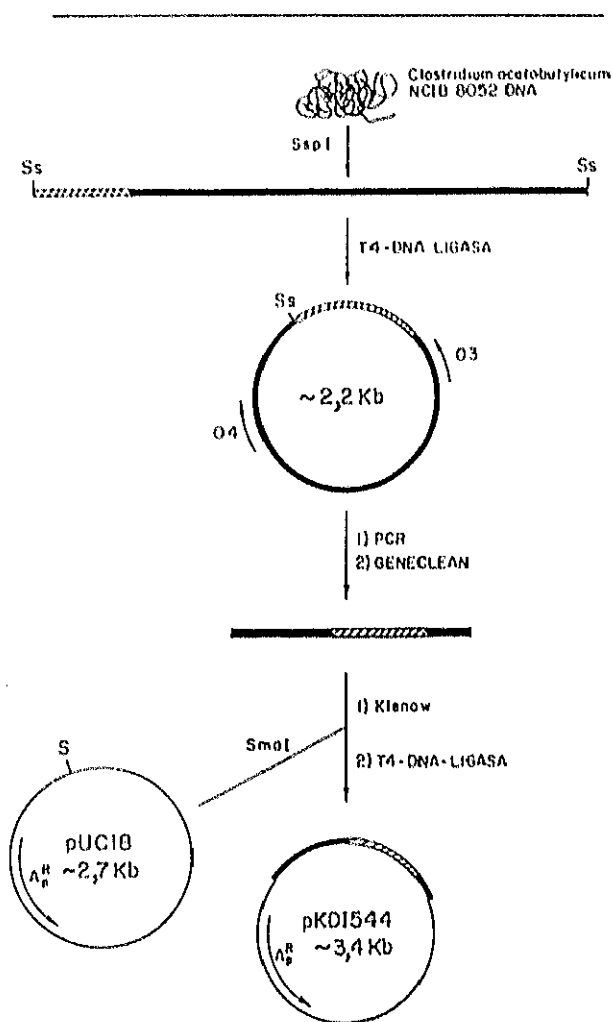


Figura 6- Estrategia de PCR utilizada para la clonación del extremo 5' terminal del gen *espA*. El trazo grueso negro corresponde a la región de DNA de secuencia conocida. El trazo rayado corresponde a la región de DNA que nos interesa clonar y secuenciar. El trazo fino corresponde al DNA del vector de clonación pUC18. O3 (5'-ATAAGCAGGCACTAAAGCCATAA-3') y O4 (5'-CTCAAACCTTCAGCAGGATGGACTT-3') son los oligonucleótidos sintéticos utilizados para la amplificación por PCR de la región 5' del gen *espA*. Las abreviaturas utilizadas: Ap^R, resistencia a Ampicilina; Ss, *SspI*; S, *SmaI*. Las condiciones utilizadas en la amplificación fueron las siguientes: 1 min a 98°C para la desnaturalización, 2 min a 63 C para la reasociación y 2,5 min a 70°C para la extensión con la enzima durante los 5 primeros ciclos. Los 30 ciclos restantes se hicieron con otras condiciones: 45 s a 96°C, 1 min a 63°C y 2,5 min a 72°C.

1.2- Análisis de la secuencia de nucleótidos del gen *cspA*

La secuenciación de los plásmidos que codifican para la proteína CspA se llevó a cabo mediante la preparación de subclones de dichos plásmidos y secuenciación de los mismos mediante la utilización de los *primers* universales obtenidos a partir de la secuencia del vector (pUC18). En algunos casos, se utilizaron oligonucleótidos sintéticos como iniciadores. La secuencia de nucleótidos del gen *cspA* y la secuencia de aminoácidos deducida de la misma se muestra en la figura 7. El codón de iniciación va precedido por un posible sitio de unión al ribosoma (RBS) situado a 11 nucleótidos del ATG. Mediante la técnica de *primer extensión* ha podido determinarse el sitio de iniciación de la transcripción situado a 29 pb del codon ATG (Fig. 8). Anterior a éste se localizan dos secuencias características de promotores de genes de procariotas reconocidas por la subunidad σ -70 *E. coli* y σ -43 de *B. subtilis* (Fig. 7). La región 5' que precede al codón de iniciación es más rica en A+T que la región que codifica para la proteína CspA, siendo la relación de un 93% y 64%, respectivamente. A 27 pb del codón de terminación TAG se localiza una secuencia palindrómica capaz de formar una estructura de bucle que presenta una energía libre teórica de -17,9 kcal/mol. Esta estructura va seguida de una secuencia rica en residuos de timina y podría actuar como un terminador de la transcripción de tipo rho-independiente (Rosenberg y Court, 1979).

1.3- Análisis de la secuencia de aminoácidos y comparación con otras proteínas del banco de datos

El análisis de la secuencia de nucleótidos revela que el gen *cspA* consta de 1.774 pb que codifican para una proteína de 590 aminoácidos con M_r de 63.740 y un punto isoeléctrico de 4,1. Dicho peso molecular es sensiblemente inferior al estimado de 110.000 tras la electroforesis en SDS-PAGE. La comparación entre el amino terminal de la proteína madura, y el deducido de la secuencia de nucleótidos, revela que la proteína es inicialmente traducida con un péptido señal de 26 aminoácidos. Este péptido señal presenta las mismas características que otros péptidos señales típicos de bacterias (Simonen y Palva, 1993).

```

1      AATATTAACTAAAAAAATATATTATATATATTTTAAAAAAGTACCCAAAGTGAATTTTGACATTGTATTATTAGTGATAT
84  TACAGTAAATA-35TACGAAGAA-10AGTAAATATRBSAAAGTAATA*ATTITATGTACAGCATTTATGTAATGGGATCATGTTATAGATTAATGTAAATGTGTAATAAATFACATGTGATT
204  TATTTGATTAC-35TAAGAAATTTATA-10AAAAATATARBSTATTA-35TAGTTATATATA-10TTTATAGATGTGTA-10AAATA-10GTATATA-10TTTAT-10GAATTTTAAATTTAGGAGGAAAAAAGTT
1*  H F E R A N K I T S L L V A A A S V H A L V F A Y A A D V K K V D S E D G T V Y
324  ATGTTTAAAGAG*AAAGAAATTAATCTTTTATTAGTAACTGCTGCTT-35AGTTAGGCTTTAGTGGCTGCTTATGCTGAGAGCTAAAGAAAGTTGATT-10AGAAGATGGTACTGTATAC
41*  N A V A Y K D G K Y F V D G E I N D D K E A Y Y V A D G K F N K L E D V D S G D
444  AATGCAITAG*ATA-35AAAGATGGTAAATACTTTGTGATGAGAAAT-10AAAGATGATGAAGAGCTTACTATGTAAGTGAAGAAAT-10AA-10AAATAGAGATGTTGATTCAGGAGAT
81*  E A V L F G E N Y L D V S D G D Y T V D L D K G S V T D D D V K G D T A D D A A
564  GAAGAGTTTATTTGGAGAAAAATCTTATGATGTATCAGAG-35AGAGACTACACTGTTGATTATAGATAAAGTAGTGTAACTGATGTAGAGTTAAAGGTGATACAGCAGAGGATGCTGCA
121*  A A L R E K K I K D D T D D R Y L E S E A E G V K N E D D L A I I G G A K Y D K P
604  GCAGCTTTAAGAAA-35AAAAATTAAGATGATA-10AGATGATAGATATTTAGAAAGTGAAGCAGAAGGCTTAAAAATGAAGATGATCTTGTCTATAATTGGAGAGCAAGATGACAAAGCTT
161*  W Y A T T Y T A S A K A I G D V N G L T A T N N K F N V Y T D T N G A Y I D A D
804  TGGTATCTACAGCATATATCTGCTTGTGCAAAAGCTATAGCTGATGTTAATGGATTAAACAGCAACAATAAATAGTTCAATGTTTATAGTATGATACAAATGGTCTTACATTGATGCAGAT
201*  Y N L G E V K V T T T A D G A T K E V T V E N T N D T Y D A A G S S D S G K D K
924  TATACTTAGGAAAGTAAAGTTTAAACAACTGCTGACGCTGCTAC-35AAAGAGTTTACTGTAGAAAACACAAACGATCTTATGATGCTGCTGCTTATAGGATTCAGGAAAGATAAA
241*  V S A S V K Q T A V L T Q D E D N I Y R L V E V T V K T G H G V I T E I N G V K
1044  GTAAGTGAAGTGT-35TAACAAATCTGAGTTTAAAGAAAGATAAGATATATCTACAGACTTGTAGAAAGTAACTGTTAAACAGGCTCATGCTGTTATACAGAAATTAATGGCTAAAA
281*  I A D I G A N T V F G G T N T E V T F P A I Q N I S K A Q A S D D V D G A K Y A
1164  ATAGTGACATTGCTGCTAATACAGTATTGCTGGTACAAAT-35CTGAAGTAAGCTTCCAGCTATCTCAAAAGATATCAAAAGCACAAGCTTCTGATGATGTTGATGGAGCTAAATATGCT
321*  K T V T T Y A L S D D S G N K L D E Q S L F V K T D G T A V T T T K Y T V V N G
1284  AAAACAGTAATCTACTTATGCTTTATCTGATGATGAGGAAATAAATAGATGAACAAAGTTTATTTGTAAGAACTGATGGTACTGCTGTTACTACAACTAAGTATATCTGTAGTCAATGGA
361*  K L I A Y N T D I N D N K K V T V D A Y T L K T N G G Y Y A D E E D K S E E D
1404  AAGCTTATAGCTTACAACTACTGATATTAAATGCAATAAAAAAGTAAGCTGTTGATGCTATATACATTAAAGACAAAAGGTGGATCTACTACCTGATGAGAAAGATAAGAGCGGAAGAGAT
401*  C E V S A Q D K T T A A V O T D V D G N L H R L D G G Y I Y K F D N T D D W D K
1524  TGGGAAGTAAGTGCTCAAGATAAAACACAGCAGCTGTTCAACAGAGCTTGTAGGAAATCTTTGGAGATTAGATGGTGTATATATCTACAAATCTGATAACACTGATGATTGGGATAAA
441*  V Y K V D G S F D E L S V Y D K D N H V A W S E D D D V Y S L I G G K S D D N K
1644  GTTTACAAAGTAGAGGGATCTTTCTGATGAATTATCAGTTTACGATAAAGACAAACATGGTTCTTGGAGCGAAGATGATGATGATATTATTAAATCGGAGGAAAGAGCGATGATAATAAG
1-> P1 1-> P2
481*  G D D Q G T T P V V K A G W A Q T S A G H T Y V K A D G T K A T G H L Q D G G A
1764  GGTGATGACCAAGGAACAACCTCTGTAGTTAAAGCTGGTTGGGCTCAAACTTCAAGCAGATGGACTTATGTTAAAGCTGATGGAACTAAAGCTACTGTGTTACAGAGCGGTGGTCT
1-> P3 1-> P4
521*  W Y Y L K A D G T H A T G W I Q D G A T W Y Y L N G S G A H Q T G H L N D N G T
1884  TGGTACTACTTAAAGCTGATGGTACAAATGGCTACAGGTTGGATTCAAGATGGAGCAACTTGGTACTACTTAAACGGATCAGGTGCTATGCAACTGGTTTAAATGATAACGGAACT
1-> T
561*  F Y Y L N G S G A H L S N T T T P D G Y Y V G A N G A H V K
2004  TTCTACTACTTAAATGGATCAGGTGCTATGTTATCTAACACAACACTCTGATGGATCTATGTAGGAGCTAATGGAGCTTGGGTTAAATAGT-35TTTAACTGAAGCTTATTAGATAAAT
2124  AAAAGGACAAGGCTTAGCTTGTCTTTTTTTTATCAATAATATTATATATAATGAATTACATTTGAAACTCTAATGGGCTAAGTGGCTCAATTAAAGATTAGAAAGTAGAAAAATCAATT
2244  GTTATATATTAGATAGGAAAAAATTTTATGTATATGAATCTTTGAAGATATAAGAAAAATAGAGTTGATCTCTCTATGTAGAGCTTTGATTATTAAATTTACTTATATTAGTATAT
2364  AATAGCTATATGAGTTTAAA

```

Figura 7- Secuencia de nucleótidos del gen *cspA*. Se muestra la secuencia de aminoácidos en código de una letra por encima de la secuencia de nucleótidos. Las dos posibles secuencias promotoras -10 y -35, el sitio de unión al ribosoma (RBS) y el posible terminador de transcripción aparecen subrayados. P1, P2, P3 y P4 hacen referencia a los motivos de unión a colina presentes en la proteína CspA y T a un posible motivo imperfecto. El codón de terminación (TAG) se indica con un asterisco.

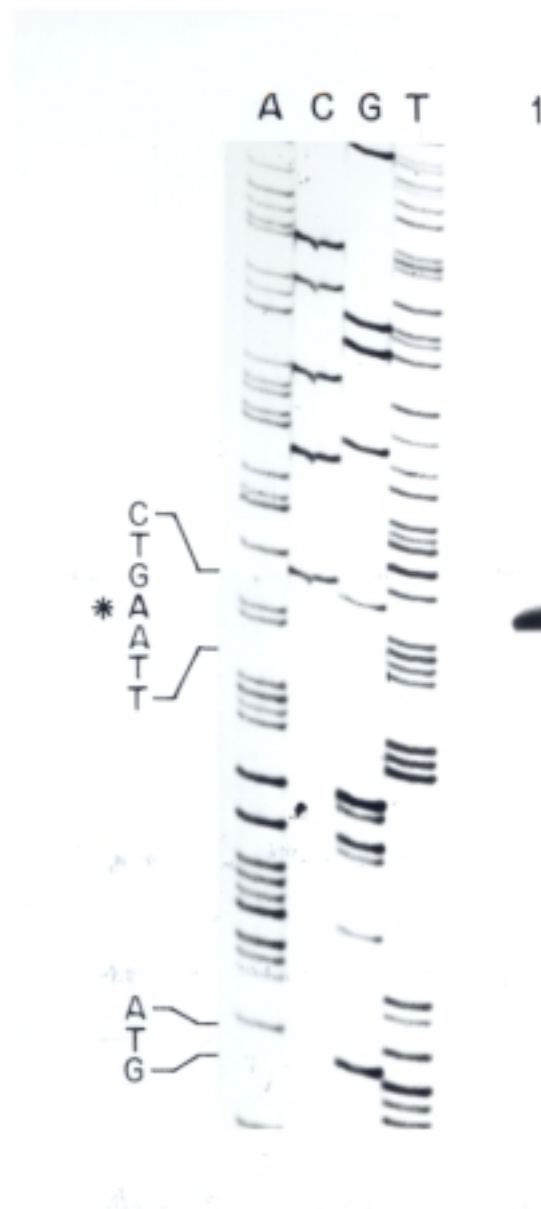


Figura 8- Análisis del sitio de iniciación de la transcripción mediante *primer extensión*. A la izquierda de la figura se muestra la secuencia del plásmido pKD1544 realizada con el oligonucleótido O3 (Fig. 5). A la derecha (1) se muestra la migración de la banda correspondiente al sitio de iniciación de la transcripción que se corresponde con la A marcada con un asterisco. Se señala también la situación del codón de iniciación de la traducción ATG.

El perfil de hidrofobicidad (HP), según describen Kyte y Doolittle (1982), indica que la mayor parte de la proteína es hidrofílica (Fig. 9). La proteína está constituida por un 18% de residuos ácidos (Glu y Asp) y un 10% de residuos básicos (Lys, Arg e His). Sólo hay presente una cisteína en toda la molécula. También se han localizado cuatro sitios potenciales de N-glicosilación (N-X-T/S) en las posiciones 545, 558, 565 y 573 (Fig. 7) (Paul y cols., 1986).

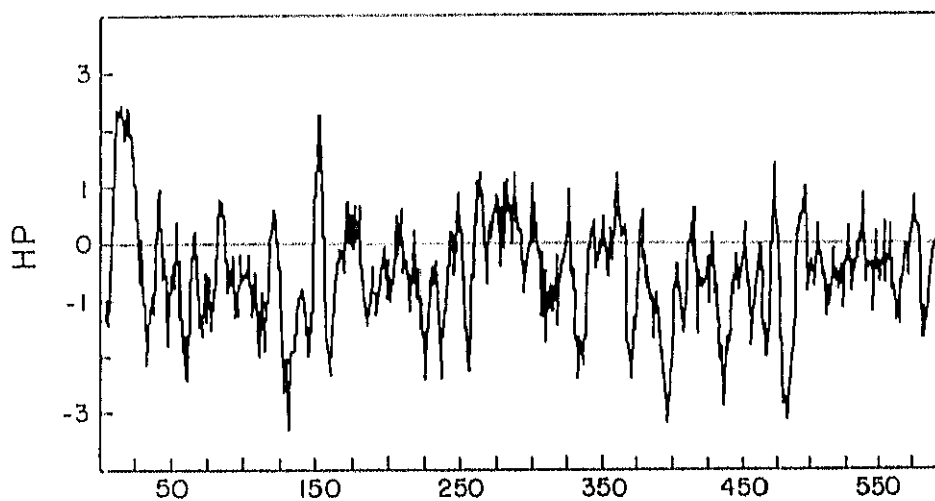


Figura 9- Perfil de hidrofobicidad de la proteína CspA. Se llevó a cabo de acuerdo con el método de Kyte y Doolittle (1982).

La comparación de la proteína CspA con las incluidas en la base de datos del EMBL reveló que la región C-terminal de esta proteína era muy similar al dominio de unión a colina de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos (García y cols., 1990; López y cols., 1992; Ronda y cols., 1987) y del antígeno PspA (Yother y cols., 1994) (Fig. 57, pag. 162). También presentaba una importante similitud con las toxinas de *C. difficile*, y con la GBP y GTFs de *Streptococcus*. Un análisis detallado de la región C-terminal mostró que estaba compuesta de cuatro motivos repetidos de 20 aminoácidos de longitud y una cola correspondiente a un quinto motivo imperfecto (Fig. 7). Dado que se ha demostrado que en las enzimas de neumococo estos módulos confieren la especificidad de unión a la colina, es evidente que los módulos de esta proteína son los responsables de la unión a

dicho ligando. Sin embargo, el dominio N-terminal no mostró una similitud considerable con las proteínas de la base de datos.

Los anticuerpos reaccionaron también con una proteína de M_r 125 000 que muestra en mucha menor cantidad junto con la CspA tras la purificación por cromatografía de afinidad.

1.4- Fraccionamiento subcelular

La proteína CspA, purificada por cromatografía en columnas de DEAE-celulosa y posterior electroelución de la misma, se utilizó para la obtención de un suero anti-CspA lo que permitió estudiar la localización subcelular de dicha proteína mediante *Western-blot* (Fig. 10).

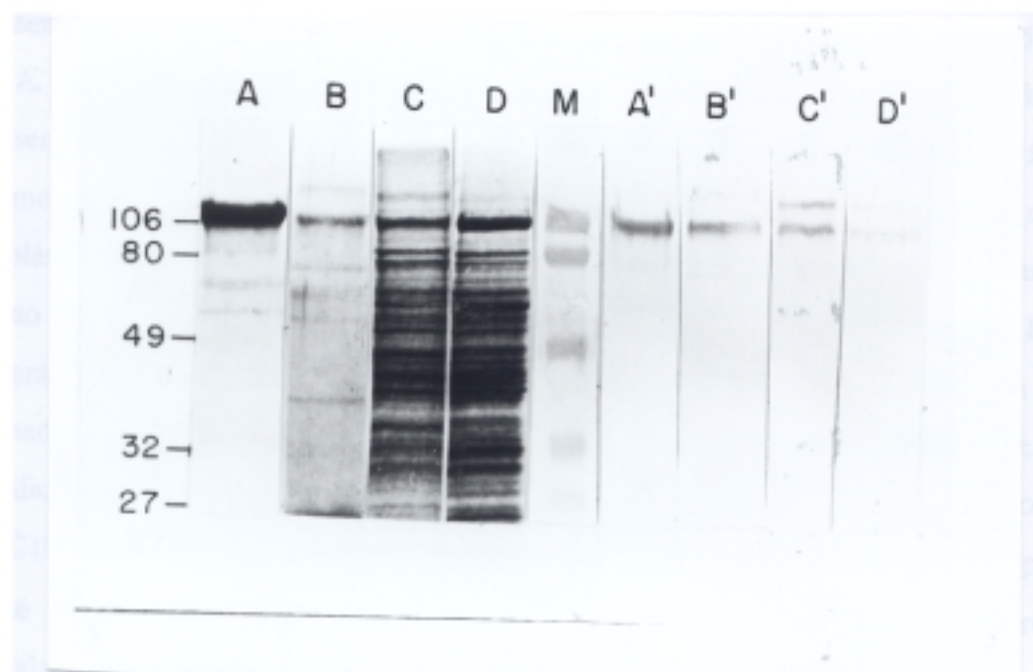


Figura 10- Distribución subcelular de la proteína CspA. Calles A y A', proteína CspA purificada mediante electroelución a partir de un extracto purificado por cromatografía de afinidad en columnas de DEAE-celulosa, a partir del sobrenadante de un cultivo de *C. acetobutylicum* NCIB 8052; Calles B y B', sobrenadante del medio de cultivo; Calles C y C', envueltas celulares; Calles D y D', fracción citoplásmica. M, marcadores de peso molecular preteñidos cuyos tamaños se muestran en la parte izquierda de la figura expresados en kDa. La parte de la figura situada a la izquierda de los marcadores corresponde a una electroforesis en SDS-PAGE de las fracciones mencionadas anteriormente y posterior tinción con azul de Coomassie. La parte de la derecha corresponde al análisis por *Western-blot* de las distintas fracciones utilizando el anticuerpo policlonal anti-CspA.

A pesar de que la proteína CspA fue inicialmente detectada en el medio de cultivo, se localiza también en la envuelta celular y, en menor cantidad, en la fracción citoplásmica. Los anticuerpos reaccionaron también con una proteína de M_r 125.000 que aparecía en mucha menor cantidad junto con la CspA tras la purificación por cromatografía de afinidad en columnas de DEAE-celulosa, sugiriendo que podría ser una proteína relacionada. Curiosamente esta proteína parece ser más abundante en la envuelta celular que en el sobrenadante del medio de cultivo.

1.5- Expresión del gen *cspA* en *E. coli*

La región estructural del gen *cspA* se reconstruyó de acuerdo con el esquema presente en la figura 11. Se utilizó un *linker* sintético que proporcionase el RBS del gen *lpp* de *E. coli*, el codón de iniciación, así como el codón correspondiente a la fenilalanina presente en la posición 2 de la proteína CspA. Esfuerzos iniciales por clonar el fragmento cromosómico *SpyI* conteniendo el gen *cspA* completo o la reconstrucción directa del gen en el plásmido pUC19 u otros vectores de expresión que presentaran promotores fuertes, como el plásmido pINIHA3 con el promotor *lpp*^p, fueron infructuosos ya que los clones obtenidos contenían siempre plásmidos delecionados. Sólomente cuando el gen *cspA* fue clonado en pUC18 en la orientación contraria al promotor *lac*, fuimos capaces de extraerlo de dicho plásmido y de clonarlo y expresarlo bajo el control de dicho promotor en el pUC19. En la figura 12A se puede observar como el crecimiento de *E. coli* (pKDI547) sufre un retraso en el crecimiento tras la inducción con IPTG (isopropil- β -D-tiogalactopiranosido) sugiriendo que la expresión del gen *cspA* en *E. coli* provoca una alteración del metabolismo del huésped.

La producción de la proteína CspA en *E. coli* (pKDI547) se detectó mediante el análisis por *Western-blot*, purificándose posteriormente mediante cromatografía de afinidad en columnas de DEAE-celulosa, dada la presencia de un dominio de unión a colina (Fig. 12B). Es interesante destacar que la proteína CspA producida en *E. coli* presenta un M_r de 80.000 el cual es superior al deducido de su secuencia de aminoácidos, pero inferior al de la proteína producida en *C. acetobutylicum*.

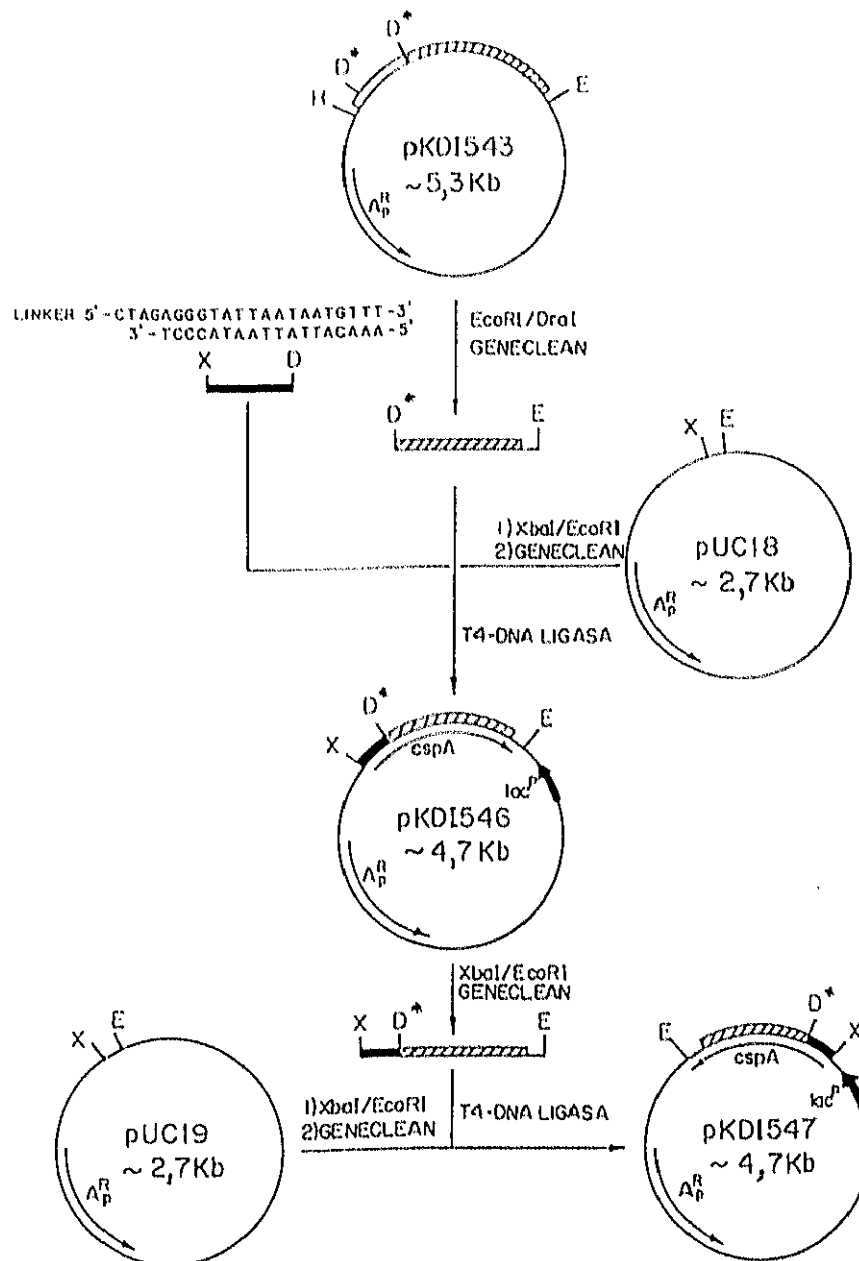


Figura 11- Reconstrucción del gen estructural *cspA*. El trazo grueso negro corresponde al *linker* utilizado para introducir los codones que codifican para los dos primeros aminoácidos del gen estructural *cspA* así como el sitio de unión a ribosoma del gen *lpp* (lipoproteína) de *E. coli*. El trazo grueso rayado corresponde al gen estructural *cspA* incompleto. El trazo fino corresponde al DNA plasmídico. Símbolos: D, *DraI*; E, *EcoRI*; X, *XbaI*; lac^P, promotor de los genes del operón de la lactosa; Ap^R, resistencia a ampicilina. Los asteriscos indican que existen otras dianas de restricción para la misma enzima en el DNA plasmídico.

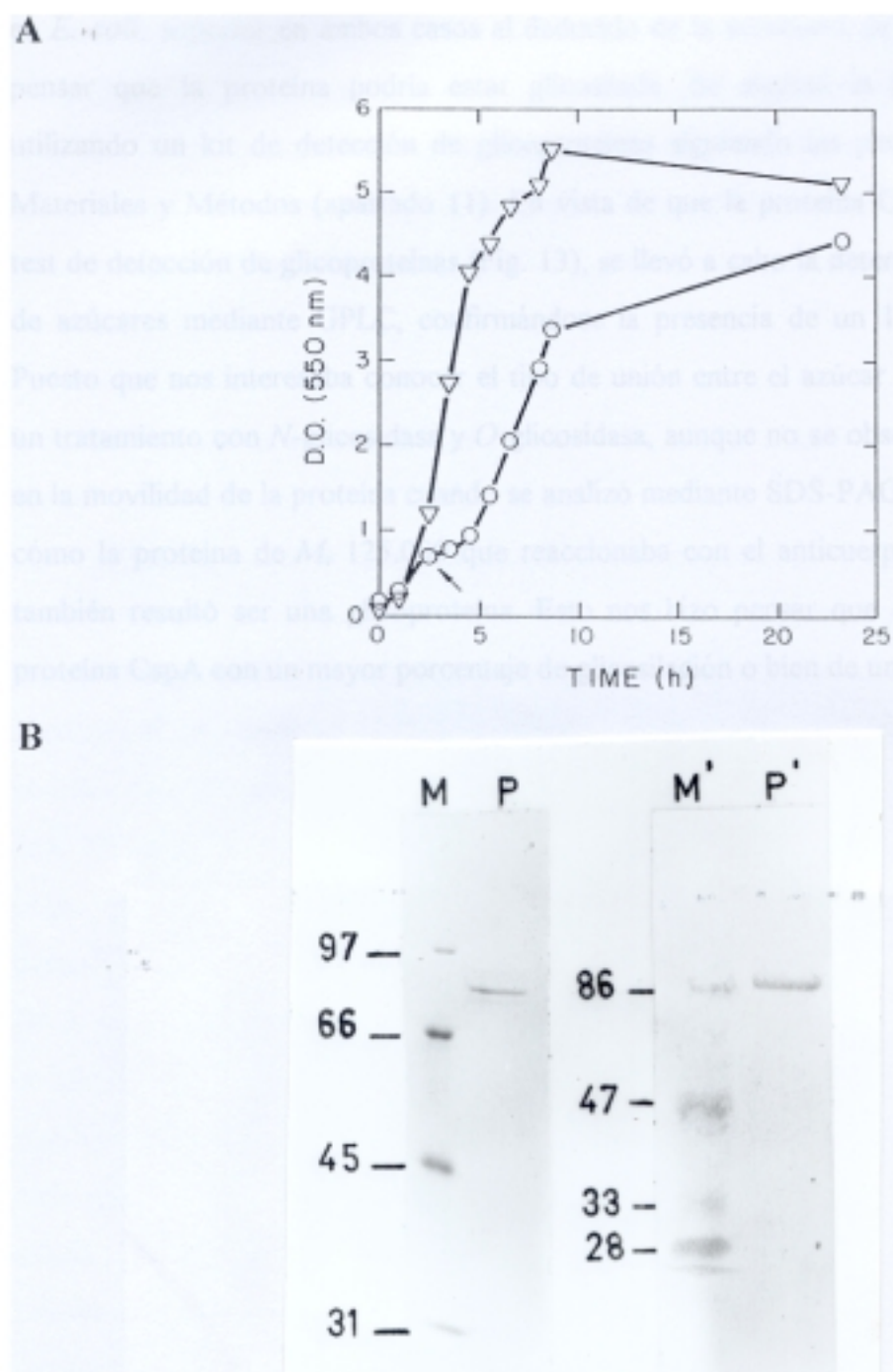


Figura 12- Expresión del gen *cspA* en *E. coli*. A) Curva de crecimiento de *E. coli* TG-1 (pUC19) (▽) y *E. coli* TG-1 (pKDI547) (○). La flecha indica el momento en que se añadió el agente inductor IPTG (1 mM) al medio de cultivo. B) El panel de la izquierda corresponde a la tinción con azul de Coomassie de la proteína CspA producida en *E. coli* y purificada en columnas de afinidad en DEAE-celulosa. El panel de la derecha corresponde al análisis mediante *Western-blot* de dicha proteína utilizando el anticuerpo policlonal dirigido contra la proteína CspA. Los marcadores de peso molecular standard (línea M) y los preteñidos (línea M') se expresan en kDa. D.O., densidad óptica.

La diferencia de tamaño entre la proteína expresada en *Clostridium* y la expresada en *E. coli*, superior en ambos casos al deducido de la secuencia de aminoácidos, nos hizo pensar que la proteína podría estar glicosilada. Se analizó la presencia de azúcares utilizando un kit de detección de glicoproteínas siguiendo las pautas que se indican en Materiales y Métodos (apartado 11). En vista de que la proteína CspA dió positivo en el test de detección de glicoproteínas (Fig. 13), se llevó a cabo la determinación del contenido de azúcares mediante GPLC, confirmandose la presencia de un 12% (p/p) de ramnosa. Puesto que nos interesaba conocer el tipo de unión entre el azúcar y la proteína se realizó un tratamiento con *N*-glicosidasa y *O*-glicosidasa, aunque no se observó ninguna alteración en la movilidad de la proteína cuando se analizó mediante SDS-PAGE. Es curioso observar cómo la proteína de M_r 125.000 que reaccionaba con el anticuerpo policlonal anti-CspA también resultó ser una glicoproteína. Esto nos hizo pensar que quizás se tratase de la proteína CspA con un mayor porcentaje de glicosilación o bien de una proteína relacionada.

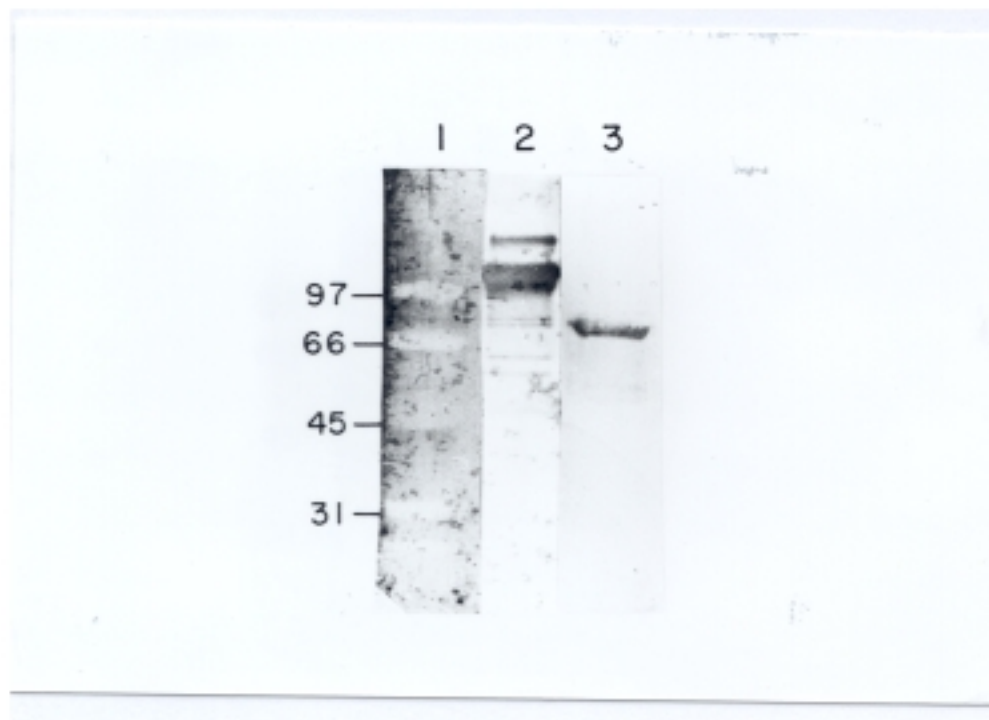


Figura 13- Detección de glicoproteínas en geles de SDS-PAGE. Calle 1, patrones de peso molecular expresados en kDa; calle 2, extracto purificado en DEAE-celulosa conteniendo la proteína CspA; calle 3, Transferrina (control positivo)

La actividad lítica de pared, utilizando extractos crudos de *E. coli* expresando la proteína CspA (pKDI547) y paredes de *C. acetobutylicum* NCIB 8052 con colina marcada radiactivamente, fue nula. Por otro lado, los anticuerpos anti-CspA no inhibieron la actividad lítica presente en el extracto purificado en DEAE-celulosa a partir del sobrenadante del cultivo de *C. acetobutylicum* NCIB 8052. Ambos resultados descartan en gran parte la posibilidad de que la proteína CspA sea una enzima lítica.

1.6- Análisis de la presencia del dominio de unión a colina en otras cepas de *C. acetobutylicum*

Para demostrar la presencia de genes homólogos al gen *cspA* en otras cepas de *C. acetobutylicum*, se realizó un *Southern-blot* con DNA cromosómico extraído de las mismas tras la digestión con la enzima de restricción *HincII*. La hibridación se realizó en condiciones no estrictas (45°C), utilizando como sonda el fragmento *PvuII-SspI* que codifica para el dominio de unión a colina de la proteína CspA (Fig. 5). Además de en la cepa NCIB 8052, las cepas NI-4 y NI-504 en las cuales se había descrito la presencia de colina en la pared (Podvin y cols., 1988), contenían secuencias que hibridaban con el dominio de unión a colina del gen *cspA* (Fig. 14). Estos resultados muestran también que el DNA cromosómico de la cepa NCIB 8052 contiene al menos cuatro secuencias que deben codificar para dominios de unión a colina. Respecto a la cepa NI-504, a pesar de estar considerada como un mutante de la NI-4, presenta un perfil de hibridación diferente. Con la cepa ATCC 824 no aparecía ninguna banda de hibridación como cabría esperar pues es sabido que no contiene colina en su pared. Además Wilkinson y cols. (1993) habían propuesto que las cepas NI-4 y NCIB 8052 eran la misma, y sin embargo nuestros resultados indican que se trata de cepas diferentes. Todos estos datos revelan la posible existencia de genes que codifican para proteínas con dominios de unión a colina en las cepas de *C. acetobutylicum* que contienen colina en la pared.

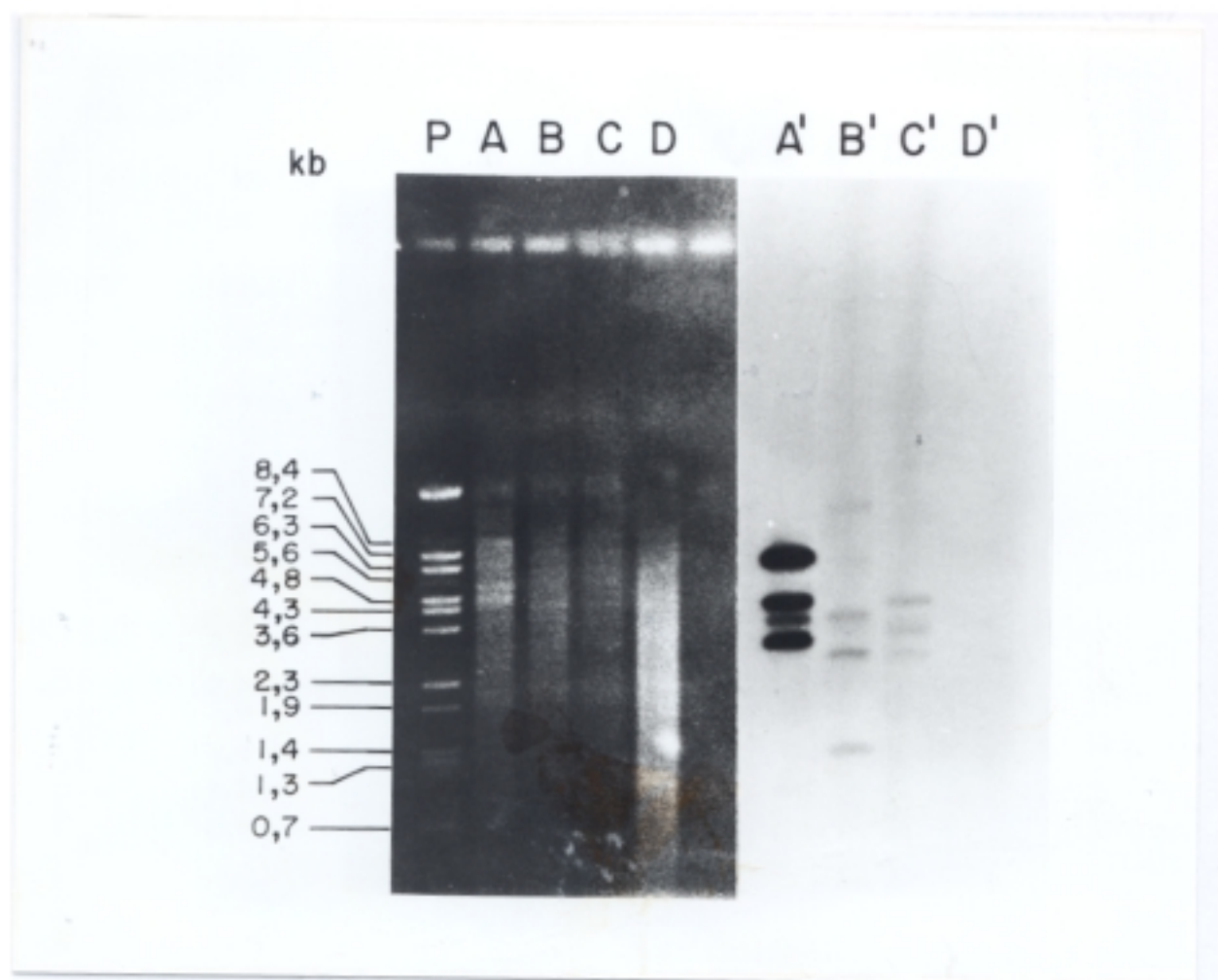


Figura 14- Análisis mediante *Southern-blot* del DNA de diferentes cepas de *C. acetobutylicum*. El panel de la izquierda muestra las digestiones *HincII* de los DNAs de diferentes cepas de *C. acetobutylicum* tras la tinción con BrEt y visualización mediante radiación UV. El panel de la derecha muestra la autorradigrafía del análisis mediante *Southern-blot*. Las cepas de *C. acetobutylicum* analizadas fueron: NCIB 8052 (calles A y A'), NI-504 (calles B y B'), NI-4 (calles C y C') y ATCC 824 (calles D y D'). Los marcadores de peso molecular (*MBstEII*) se muestran a la izquierda en kb. La sonda utilizada fue el fragmento *PvuII-SspI* que codifica para el dominio de unión a colina del gen *cspA*. La temperatura de hibridación y lavado fue de 45°C

2- CARACTERIZACIÓN MOLECULAR DE UNA FAMILIA DE GENES QUE CODIFICAN PARA PROTEÍNAS CON DOMINIOS DE UNIÓN A COLINA (Csp)

secuencia nucleotídica del mismo. El primer fragmento...

La hibridación del DNA de *C. acetobutylicum* NCIB 8052 digerido con *HincII* con el fragmento del gen que codifica para el dominio de unión a colina de la proteína CspA, daba lugar a la aparición de cuatro bandas, correspondiendo la más intensa a la del gen *cspA* (Fig. 14). Se estudió la posibilidad de clonar los genes correspondientes a las otras tres bandas de hibridación puesto que, al igual que en neumococo, podrían codificar para otras proteínas implicadas en el metabolismo de la pared bacteriana. La selección de las enzimas de restricción más adecuadas para la construcción de genotecas, se realizó en base a un experimento de *Southern-blot* con el DNA de *C. acetobutylicum* NCIB 8052 digerido con diferentes enzimas de restricción e hibridado a 55 °C con la sonda que codifica para el dominio de unión a colina del gen *cspA* (Figura 15). De esta manera se clonaron a partir de una genoteca *EcoRV* dos fragmentos correspondientes a dos de las bandas de hibridación y mediante una genoteca *SspI* el tercero. A continuación se describe la estrategia de clonación de estos tres genes así como su secuenciación.

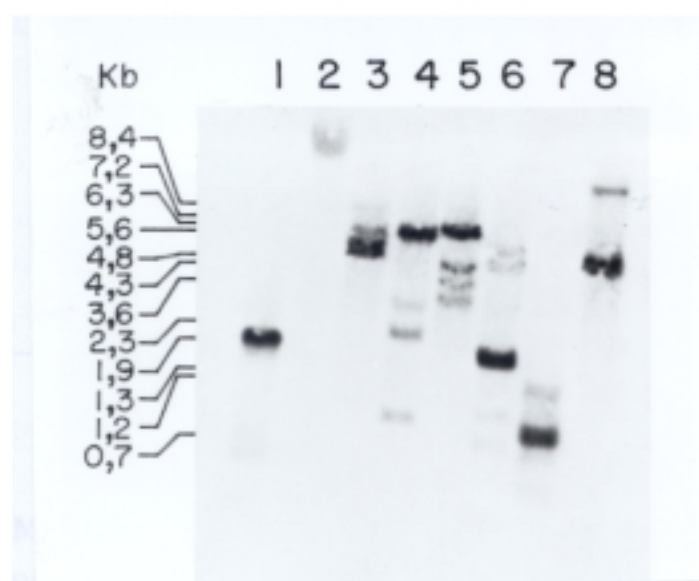


Figura 15- Análisis mediante *Southern-blot* del DNA de *C. acetobutylicum* NCIB 8052. El DNA se digirió con las enzimas *DraI* (calle 1), *SspI* (calle 2), *EcoRI* (calle 3), *EcoRV* (calle 4), *HincII* (calle 5), *HindIII* (calle 6), *PstI* (calle 7) y *XbaI* (calle 8). En el margen izquierdo de la figura se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (λ /BstEII). Se utilizó como sonda el fragmento *PvuII-SspI* del plásmido pKDE543 que codifica para el dominio de unión a colina de la proteína CspA. La temperatura de hibridación y lavado fue de 55°C.

2.1- Clonación y secuenciación del gen *cspB*

Este gen ha sido clonado en tres fragmentos que nos han permitido determinar la secuencia nucleotídica del mismo. El primer fragmento clonado fue obtenido mediante el análisis de una genoteca *EcoRV* del DNA de *C. acetobutylicum* NCIB 8052 preparada en pUC18 digerido con *SmaI* y utilizando como huésped *E. coli* TG-1. Las genotecas se analizaron mediante hibridación de DNA sobre filtros de colonias celulares en condiciones no estrictas (55°C), utilizando como sonda el fragmento *PvuII-SspI* del plásmido pKDE543, que codifica para el dominio de unión a colina de la proteína CspA. De esta manera se obtuvo un clon positivo que contenía el plásmido recombinante pKC33, con un inserto de aproximadamente 3 kb (Fig. 16), que contenía la región que codificaba para el dominio de unión a colina de dicha proteína, tal como esperábamos tras los experimentos de *Southern-blot* realizados previamente.

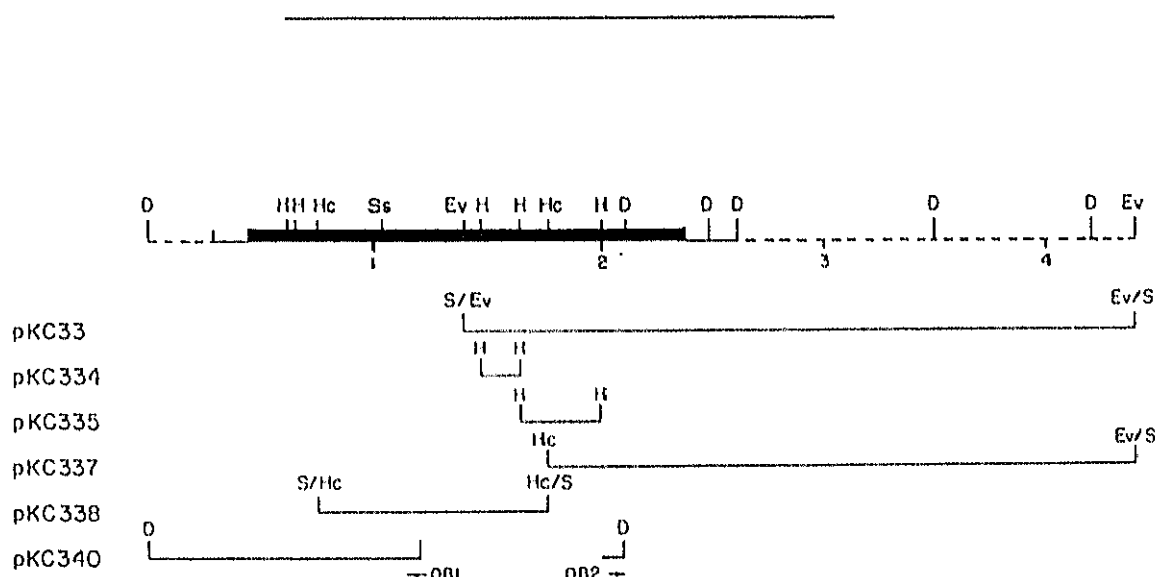


Figura 16- Mapa físico del gen *cspB*. En la parte superior de la figura se muestra el mapa físico del fragmento de DNA que contiene la secuencia completa del gen *cspB*, resultado de la unión de los insertos de los plásmidos pKC33, pKC338 y pKC340. El trazo negro grueso hace referencia al gen estructural *cspB*. El trazo continuo corresponde al fragmento de DNA secuenciado (Véase Fig. 18). El trazo discontinuo es DNA que queda por secuenciar. Los trazos finos de la parte inferior de la figura hacen referencia a los tamaños de los insertos clonados y utilizados en la secuenciación del gen *cspB* y regiones no codificantes. OB1 (5'-CGGAATTCCGTCAGCTGCATCGTAAGCGTCG-3') y OB2 (CGGAATTCCGGCGACAGGTT GGG-3') son los oligonucleótidos sintéticos utilizados para la clonación mediante IPCR del extremo 5' del gen *cspB*. Dianas de restricción: Ev, *EcoRV*; H, *HindIII*; Hc, *HincII*; D, *DraI*; Ss, *SspI*; S, *SmaI*.

En un intento por clonar la región 5' del gen, se analizó una genoteca *HincII* construida en pUC18 digerido con *SmaI*, utilizando como sonda el fragmento *HincII* de 38^{nt} pb del plásmido pKC33 (Fig. 16), basándonos en los datos previos obtenidos mediante un experimento de *Southern-blot* hibridado con dicha sonda (Figura 17).

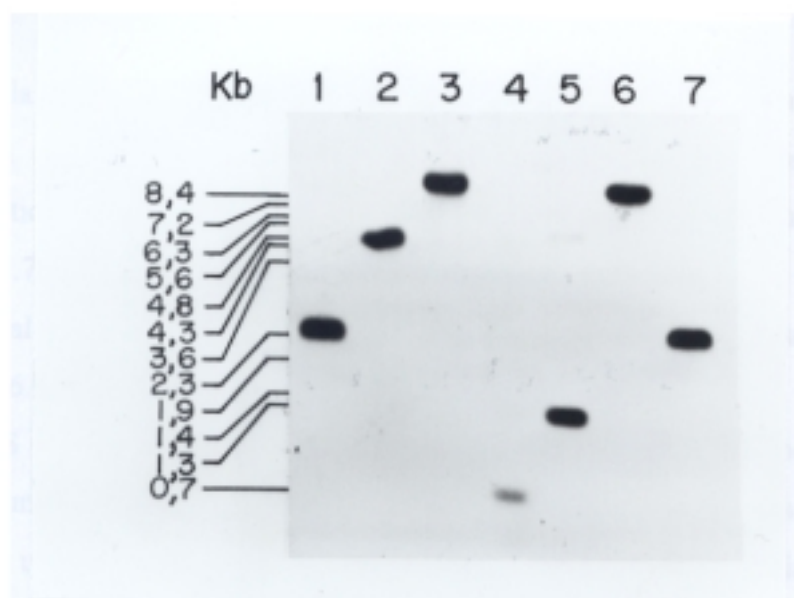


Figura 17- Análisis mediante *Southern-blot* del DNA de *C. acetobutylicum* NCIB 8052.

El DNA se digirió con las enzimas *SspI* (calle 1), *PvuII* (calle 2), *PstI* (calle 3), *HindIII* (calle 4), *HincII* (calle 5), *EcoRI* (calle 6), *DraI* (calle 7). En el margen izquierdo de la figura se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (*λBstEII*). Se utilizó como sonda el fragmento *HincII* del plásmido pKC33. La temperatura de hibridación y lavado fue de 65°C.

De esta manera se aisló el plásmido recombinante pKC338 con un inserto de 1 kb (Fig. 16). Como aún estaba incompleta la secuencia del gen *cspB*, se siguió la técnica del IPCR al igual que en el caso del gen *cspA*, para terminar de clonar lo que nos faltaba. La enzima de elección, en este caso, fue *DraI* que da lugar a una banda de restricción superior en tamaño a la de *HincII* y, por tanto, con una mayor probabilidad de clonar el resto de la secuencia del gen (Fig. 17). Para ello la estrategia seguida fue igual a la que se muestra en la figura 6, pero en este caso el DNA cromosómico se digirió con *DraI*. El DNA recircularizado se utilizó como molde para la amplificación con los oligonucleótidos OB1 y OB2 del fragmento que nos interesaba (Fig. 16), siendo las condiciones del PCR las mismas

que en el caso del gen *cspA* pero con una temperatura de anillamiento de 50°C. Puesto que los oligonucleótidos fueron diseñados con un sitio de restricción *EcoRI*, el fragmento amplificado de 1,1 kb y el vector de clonación pUC18 se digirieron con dicha enzima, se ligaron y con la mezcla de ligación se transformó *E. coli* TG-1. Mediante hibridación de colonias, utilizando como sonda radiactiva el fragmento *BamHI-SspI* de 278 pb del plásmido pKC338, se aisló un clon positivo con el plásmido recombinante pKC340 (Fig. 16).

En la figura 18 aparece la secuencia completa del gen realizada mediante la preparación de subclones de dichos plásmidos y secuenciación de los mismos con oligonucleótidos sintéticos y *primers* universales. Se puede distinguir un marco de lectura abierto de 1.708 pb, que codifica para una proteína de 635 aminoácidos, correspondiendo el péptido señal a los 29 primeros residuos. La proteína sin procesar presenta un M_r 69.771, siendo de 66.959 para la proteína madura. El codón de iniciación ATG va precedido de un posible RBS (AGGAGG) a 8 nucleótidos de distancia. A 133 pb del codón de terminación se localiza una posible estructura en bucle con una energía libre teórica de -25,2 kcal/mol, seguida de una secuencia rica en residuos de timina, la cual podría actuar como un terminador de la transcripción rho independiente. Un análisis detallado de la región C-terminal indica que está compuesta por 6 motivos repetidos, estructurados y con las mismas características que los motivos de unión a colina de otras proteínas (Fig. 57, pag. 162).

2.2- Clonación y secuenciación del gen *cspC*.

El gen *cspC* se ha clonado en dos fragmentos que han permitido determinar la secuencia nucleotídica de la región que codifica para la proteína CspC madura, ya que la parte 5' del gen con la secuencia correspondiente al péptido señal con el ATG, el RBS y la posible región promotora, no se han podido clonar. En primer lugar se clonó un fragmento de 1,2 kb tras el análisis de un genoteca *EcoRV* construida en *E.coli* TG-1 y utilizando como vector de clonación pUC18 digerido con *SmaI*. El plásmido recombinante obtenido denominado pKC31 (Fig. 19) contenía, como se comprobó posteriormente por secuencia, una región que codificaba para el dominio de unión a colina de la proteína CspC tal y como cabría esperar al haber utilizado como sonda la región que codifica para el dominio de unión a colina del gen *cspA*.

```

1  GGTATCTATTTTAAAGAGTTGTATATTATATGTTAATGTACAAAATCTTACCTGTATATTAATAATTATTACATTTAATIGATTATTCATTATTAACATTATTAATTTTAATTAG

1'
RBS      H I R G H G K V T S L L V A A A T V A S L V P F S
121  CTTAAATATATTAAATTTTATCTTTATAGAGAGGCAATATTATGATAGAGGATAGGAAAGGTAACATCACATTAGTAGACAGACAGACAGTTCTTCTTATTAGTACCTTTTAGT

26'  G A N A A K V K R I S S D D G T I Y N A I A Y K D G K A Y I D G E I N D D E E A
241  GGGCTTAATGCTGCTGAGGTGAAAAGAAATTAATTTCTGATGATGAACTATTTATAATGCAATAGCATATAAAGACGGTAAAGCTTATATGATGGAGAAATTAATGATGACGAGAAAGCT

66'  Y Y L S N G K F N K L D D V D S G D S A A L F G E N Y L D I S D G D Y T V D L D
361  TACTATTTATCAACGGAAATTCATAAATTAGACGATGTTGATTAGGAGATAGCGCAGCTCTATTGGAGAAAAATATTAGATATATCAGATGGAGACTATAGCTGTTGACTTAGAT

106' K G S V T D D D I K G D T E D D A A A A L R K K I K D D T D D R Y N E T E A N T
481  AAAGGTAGCGTAACCTGATGATGATATTAAGGTGATCTGAAGATATGCGCGCTGCTTTAAGAAAAAAATCAAGATGATACAGATGATAGGTATAATGAACTGAAGCAAACTACT

146' I K D S N H G D L F D L I P G A K Y N K V H Y Y T Q Y K A A Q K S I D K N V N G
601  ATTAAGATCAAAATCATGGTGACTATTTGATTTAATTCAGGAGCAAGGTATAAAGGTTGGTATTATACACAATATAAGCGGCTCAAAAATCTATAGATAAAAATGTCATGGA

186' L N G L D S A H Q I E N V F T D D K G H Y I D A D Y N L G K V K V T T T A S S A
721  TTAATGGATTGGATTCAAGGACATCAATATTTAATGTATTACTGATGATAAGGCAATTATATTGATGGGACTATAACTTAGGTAAAGTTAAGGTTACACAACAGCTTCAAGTGCT

226' S G T T L T K T D Y I E N T N D A Y D A A D G I I N G T S V S G S D K L S A S V
841  TCTGGTACTACATTAAACAAACAGAGATCAATAGAGAACATAAACGAGCTTACGATGACGATGATGGAATTATCAATGGAACTAGCGTTAGCGGCTCTGATAAATTAAGTGCAGTGTT

266  V Q D R V L T Q D K D Y I Y R L A T V K I T I T T G A A A T I S E I N G V K V D
961  GTTCAAGATAGAGATTAAACCAAGATAAAGATTATATATATAGACTTCAACTGTAAAAAATACTATTACCCTGGTGACGCTGCTACAATTAGTGAATAAATGGTGTAAAGTTGAT

306' P H N S N D I F K V E N N G Q V V S F K A I Q K I S K T Q A S G D I D D A K Y A
1081 CCAATAATAGCAACGATATCTTCAAAAGTTGAAAATAACGGCAAGTTGTATCTTTCAAAAGCTATTCAAAAATATCAAAAACCAAGCTTCTGGTGATATTGATGACCGCTAAATATGCT

346  K T V T T Y A L S D K D G K K L D A E D L F I N T S G H V V T T T N Y T V G S G
1201 AAACTGTAACTACTATATGACTTTCTGTAGAAAGATGATAAAGTTAGATGCTGAGGATTATTATATAAATCTTACGGCAAGCTAGTCACACAACCTAATCTACTGTAGTTCTGGA

386' K L I A Y N S E I H N N D K V T V R A Y T T L K S S S G F Y Y A D E E D Q S K E D
1321 AAGCTTATAGCTTACAATCTGAAATTAACAATAACGACAAGTTACTGTTAGAGCTTATACATTAATAAATCAAGTAGTGGATTCTATTATGCTGATGAGGAAGATCAAGACAAAGAGAT

426' C E N S K N O G A A V Q T D V D G N L W R L D G G Y I Y K F D N T D D W D K V Y
1441 TGTGAAAACAGCAAAACCAAGGTGCTGCTGTTCAACAGATGTTGACGGAAATCTATGGAGATTAGATGGTGGATATATCTACAAATTTGATAAATCTGATGACTGGGACAAAGTATAC

466' R V D G S F D E F S V Y D K D N I V A N S K D D D V Y S L I G G K Q S N T D P D
1561 AGAGTAGACGGATCATTTGATGAATTTCTAGTTTATGATAAGATAATATCTGTTGCTTGGAGCAAGATGATGATTTTATTCTAATAAGGTGGAAAACATTAATACTGACCTGAT
1--> P1 1--> P2

506' D T P V V K T G W V Q A S D G S W S Y I K A D G N K A T G W V Q D G S T W Y Y F
1681 GATACCTCAGTAGTAAAAACAGGATGGGTACAAAGCTTCAGATGGAACCTTGGCTTTATATTAAGCTGACGGAAATAAGGCGACAGGTGGGTTCAAGATGGATCAACTTGGTACTATTTT
1--> P3 1--> P4

546' K A D G S H A T G W V Q D G S T W Y Y F Q S W G G H Q T G W L N N N G T W Y Y L
1801 AAAGCTGATGGATCTATGGCTACTGGTTGGGTTCAAGATGGATCAACTTGGTATTACTTCAATCATGGGGTGGATGCAAACTGGTTGGCTAAATAATAACGGTACTTGGTATTACTTA
1--> P5 1--> P6

586' Q S W G G H K T G W L N D N G N W Y Y L Q S W G G H Q T G W F N D N G T W Y Y L
1921 CAATCATGGGGCGAATGAAGACTGGTTGGCTAAATGATAACGGCAATGGTACTACTTACATCATGGGGCGAATGCAAACTGGTTGGTTTAAATGATAATGGTACTTGGTACTACCTA

626' Y S N G V H A A N T *
2041 TACTCAATGGTGTATGGCCCTAACAGTAGTTATGGATACACTTATCTGCAAGCGGTGCTGGGTATAAATTATCATATATAAACTATAAAATTAATATAAAATCTATATATTA

2161 ACCTAAATATATGCTCTTTTAAACCTCTCTAGGTTTACTATAAAGGCTGTCTCAAAAGAAATAGATACTTTTGGAGACAGCTTACATTTTAAATATAACGGGATCAAAATCTATAA

2281 TCTAATATCAATTTGCTTTACATAAAAGTAATATAATCTTTAAA

```

Figura 18- Secuencia de nucleótidos del gen *cspB*. Se muestra la secuencia de aminoácidos en código de una letra por encima de la secuencia de nucleótidos. El sitio de unión al ribosoma (RBS) y el posible terminador de transcripción aparecen subrayados. P1, P2, P3, P4, P5 y P6 hacen referencia a los motivos de unión a colina presentes en la proteína CspB. El codón de terminación (TAG) se indica con un asterisco.

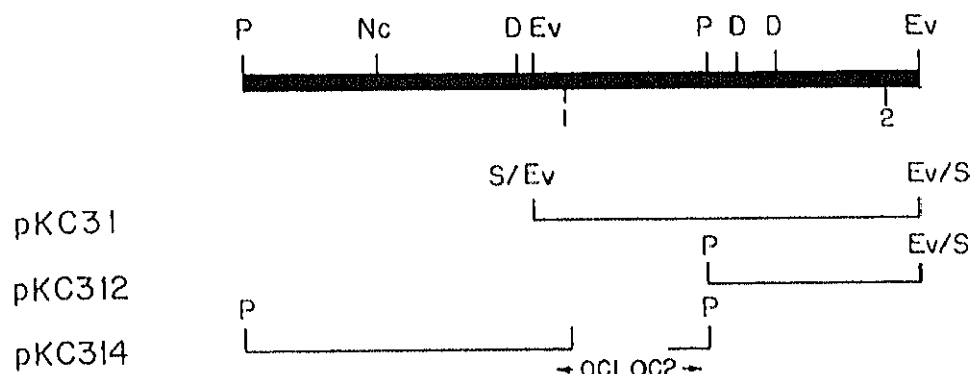


Figura 19- Mapa físico del gen estructural *cspC*. En la parte superior de la figura se muestra el mapa físico del fragmento de DNA que contienen la secuencia incompleta del gen *cspC* resultado de la unión de los insertos de los plásmidos pKC31 y pKC314. Los trazos finos de la parte inferior de la figura hacen referencia a los tamaños de los insertos clonados y que fueron utilizados en la secuenciación del gen *cspC* y regiones no codificantes. OC1 (5'-GCTGGATCAACATCATCGGAGTC-3') y OC2 (5'-CGATGGATCTTTTGATGA GCTTTC-3') son los oligonucleótidos sintéticos utilizados para la clonación mediante IPCR del extremo 5' del gen *cspC*. Dianas de restricción: Ev, *EcoRV*; P, *PstI*; D, *DraI*; S, *SmaI*; N, *NcoI*.

Puesto que el gen estaba incompleto se procedió de nuevo a la clonación de la región 5' del gen mediante IPCR. La enzima de restricción *PstI*, se escogió tras la realización del experimento de *Southern-blot* que se muestra en la figura 20. La estrategia seguida fue la misma que en casos anteriores al igual que las condiciones utilizadas para el PCR, aunque con una temperatura de anillamiento de 60°C. En este caso, los oligonucleótidos OC1 y OC2 utilizados para el PCR no fueron diseñados con ninguna diana de restricción por lo que el fragmento amplificado de 1,1 kb, después de tratarlo con el fragmento Klenow de la DNA polimerasa, se clonó en un pUC18 digerido con *SmaI*, empleando como cepa receptora *E. coli* TG-1. Mediante hibridación de colonias, utilizando como sonda radiactiva el fragmento *PstI* de 551 pb del plásmido pKC31, se aisló un clon positivo con el plásmido recombinante pKC314 que contiene la mayor parte de la región 5' del gen estructural a excepción de la región que codifica para el péptido señal (Fig. 19).

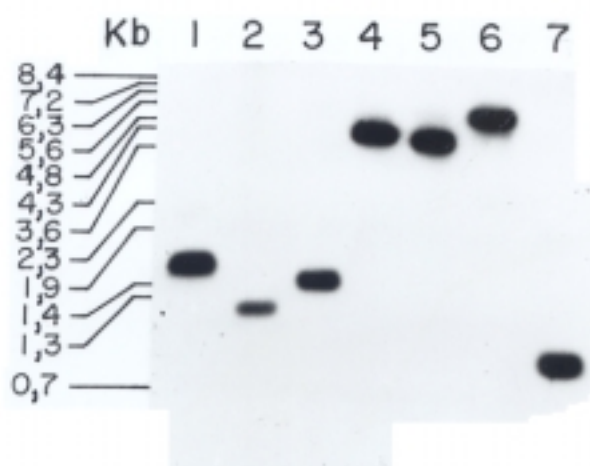


Figura 20- Análisis mediante *Southern-blot* del DNA de *C. acetobutylicum* NCIB 8052.

El DNA se digirió con las enzimas *SspI* (calle 1), *PvuII* (calle 2), *PstI* (calle 3), *HindIII* (calle 4), *HincII* (calle 5), *EcoRI* (calle 6), *DraI* (calle 7). En el margen izquierdo de la figura se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (*λBstEII*). Se utilizó como sonda el fragmento *HincII* del plásmido pKC31. La temperatura de hibridación y lavado fue de 65°C.

En la figura 21 aparece la secuencia, en nucleótidos y aminoácidos, de la región estructural del gen que codifica para la proteína CspC procesada. La secuenciación se consiguió mediante la preparación de subclones de los plásmidos ya obtenidos y la utilización de oligonucleótidos sintéticos y *primers* universales. Se puede distinguir un marco de lectura abierto constituido por 1.808 pb que codifican para una proteína procesada de 601 aminoácidos con un M_r de 66.832. A 29 pb del codon de terminación se localiza una posible estructura en bucle con una energía libre de -29,3 kcal/mol, seguida de una cola de poli-T, pudiendo actuar también como un posible terminador de la transcripción de tipo rho independiente. La proteína CspC presenta también un dominio de unión a colina constituido por 5 motivos repetidos (Fig. 57, pag 162) y una cola terminal de 17 aminoácidos (Fig. 21).

2.3- Clonación y secuenciación del gen *CspD*.

El gen *cspD* se clonó en dos fragmentos aislados, en ambos casos mediante el análisis de genotecas. En primer lugar, se clonó un fragmento de 1,6 kb en pUC18 digerido con *SmaI* tras el chequeo de un genoteca *SspI* construida en *E. coli* TG-1. El clon positivo

1 CT

1' A D V K R I D S D D G T V Y N A V A Y K D G K F F V D G E I N D D E E A Y Y V S
3 G C A G A T G T A A A A A A A T T G A T T C A G A T G A T G G T A C T G T A T A C A A T G A G T A G C T T A C A A A G A T G T A A A T C T T T G T T G A T G G A G A A A T T A A C G A T G A C G A A G A A G C C T A C T A T G T A T C T

41' D G K F N K L E D I D S G D D A D L F G E K Y L D I N D G D Y T V D L D K G S I
123 G A T G R A A A A T T A A C A A A C T A G A A G A T A T T G A T T C A G G A G A T G A C G C T G A T C T A T T C G G A G A A A A T A T C T T G A T A T A A A T G A T G G T G A T T A C A C T G T T G A T T A G A T A A A G G T A G T A T A

81' T D D D T K G D T A D D A S A D L R K K I K D D T D D R Y D E D F A D N I I D E
243 A C T G A T G A T G A T A C T A A G G G G A T A C T G C T G A T G A C G C T T C A G C A G A T C T A A G A A A A A A T T A A A G A T G A C A C T G A T G A C A G A T A T G A A G A T T T T G C T G A T A T A T A A T A G A T G A A

121' N K D D T E D L T K L N I I P G A K Y K T P W Y Y T Q Y K A A D K A V N D N V N
363 A A T A A G A T G A T A C T G A A G A T T T A A C T A A A T A A T A A T T C C A G G T G C T A A A T A T A A A C T C C A T G G T A T T A T A C A C A A T A T A A A G C A G A T A A A G C T G T G A A T G A T A A T G T T A A T

161' G L K E I T D Q N K R V F N V Y T D T K G G Y I D A D Y N L G K V K V K T T A D
483 G G A C T A A A G A A A T A A C T G A T C A A A A T A A A C G A G T T T C A A T G T T A T A C T G A C A C A A A A G T G C T A C A T T G A T G C T G A C T A T A A C T T A G G A A A G T T A A G G T T A A A C C A C T G C T G A T

201' G A S T K T E T I E N T N D K Y D A A G K D D A L T A S V S Q D K I L T Q D D D
603 G G T G C A T C A A C T A A A C A G A A A C A T A G A A A A T A C T A A T G A T A A A T A C G A T C T G C T G G T A A G G A T G A T G C A C T A A C T G C A A G T T T C C C A A G A T A A A T A T T G A C A C A A G A T G A T

241' Y I Y R L A T I K V N V T T D A G A K V R I S E I N G V E L N P K N A N D K P Q
723 T A T A T T T A G A C T T G C G A C T A T A A A A G T A A A T G T A A C T A C T G A T G C T G G T G C T A A A G T T A A G T A A A T A A G T G A A A A A T G G T G T T G A C T T A A C C A A A A A C G C A A T G A C A A C C A C A A

281' I F K V E N D G G T V S F K A I Q K I S K A Q D S D D V D G A K Y A K T V T T Y
843 A T A T T T A A A G T T G A A A A T G A T G G A G A A C T G T A C A T T C A A G G C T A T T C A A A G A T A T C C A A A G C T C A A G A C T C C G A T G A T G T T G A T G G A G C T A A A T A T G C A A A A C T G T A A C T A C T A T

321' A L S D E D G K N L D A K D L F L N T D G T I N E D T K F T V A N G K L I A Y N
963 G C A C T T T C T G A T G A A G A T G G T A A A A C C T A G A T G C T A A A G A C T T A T T C C T A A A T A C T A G C G A C T A T C A A T G A A G A T A C T A A A T T A C T G T A G C T A A C G A A A A C T T A T A G C C T A C A A T

361' T E H N E Y K K V V V R A Y S L K T K G G Y Y Y A D E E D E S K E D C E V S S Q
1083 A C A G A A A T G A A T G A T A T A A A A A G T T G T T G T C A G A C A T A T T C A T T A A A A A C A A A A G G T G G A T A C T A C T A C C G C A C G A A G A A G A T G A A A G T A A A A A G A T T G T G A A G T A A G T T C T C A A

401' D D K T S A V Q T D V D G N L W R L D G G Y I Y K F D N T D D W D K V Y K V D G
1203 G A T G A T A A G A C A T C A G C T G T T C A A C A G A C G T T G A T G G A A C C T A T G G A G A T T A G A T G G T G G A T A T A T C T A T A A A T T T G A T A A T A C A G A T G A T T G G G A C A A A G T T T A C A A A G C G A T G G A

441' S F D E L S V Y D K D N I A A W S E D D D V Y S L I G G K S S D D N N T H P P T
1323 T C T T T T G A T G A G C T T T C A G T C T A T G A T A A A G A C A A T A T A G C T G C T T G G A G T G A A G A T G A T G T T A T T C A T T A A T A G G A G G A A A T C T T C T G A T G A T A A A C A C T A A T C C T C C A A C T
I--> P1 I--> P2

481' T T A G W V Q A A D G T W T Y N K E D G T K A T G W L N L N G V W Y Y L N T D G
1443 A C A A C T G C A G G A T G G G T A C A A G C T G C T G A T G G A A C T T G G A C T T A T A A T A A G A A G A T G G A A C T A A G G C T A C T G G C T G G T T A A A C T T A A A T G G A G T T T G G T A T T A T T A A A C A C T G A T G G T
I--> P3 I--> P4

521' V H A T G W L N L N G V W Y Y L N P S G A H A T S W L N L N G T W Y Y L N S S G
1563 G T T A T G C C T A C T G G T T G G T T A A A C T T A A A T G G A G T T T G G T A T T A C T T A A A T C C A C A G G A C T A T G C C T A G C T G G T T A A A C T T A A A T G G A A C T T G G T A C T A T T T A A A T C A T C A G G A
I--> P5 I--> T

561' A H A T G W L H D N G T W Y Y L N S S G A H L Y N T T V D G Y V L G S N G A H I
1683 G C T A T G C C T A C T G G T T G G T T A A A T G A T A A T G G A A C T T G G T A T T A T T A A A T C A T C A G G G C T A T G T T A T A T A C A C A A C T G T T G A T G G T T A T G T A T T A G G C T C T A A T G G A C T T G G A T T

601' R *
1803 A G A T A G T T T T A A T C T A A T T A T T C C T T T A T A A A A G G A C A A G T T T C T A T A C T G T G C T T A T A C A T C C T T A T C T A A A C G A T G G T A T A G C T T A C C T T G T T A T T T T T A T G C T A T A A C A A

1923 T A A C A T T C T A T A T C T A A A T T G A T C C C T A A C T G C A T A T A T T C A T A C T T C T A A T T T T C T A T T C C T A T A T A A C A C T A T T T T C A C C T T A A A T G T T T A T T G A T C T C A G A A C T C A T A A C

2043 C T A T T C T A T A T T T C T A A T T A A T A T C A T T T A T A A A C T C A C A C T A G A T T T T A T A T A C A G A T A T C

Figura 21- Secuencia de nucleótidos y aminoácidos del gen *cspC*. Se muestra la secuencia de aminoácidos en código de una letra por encima de la secuencia de nucleótidos. El sitio de unión al ribosoma (RBS) y el posible terminador de transcripción aparecen subrayados. P1, P2, P3, P4, y P5 hacen referencia a los motivos de unión a colina presentes en la proteína CspC y T a un posible motivo imperfecto. El codón de terminación (TAG) se indica con un asterisco.

con el plásmido recombinante pKC21 (Fig. 22), se detectó por hibridación de colonias, utilizando como sonda el fragmento *PvuII-SspI* que contiene la secuencia del dominio de unión a colina del gen *cspA*.

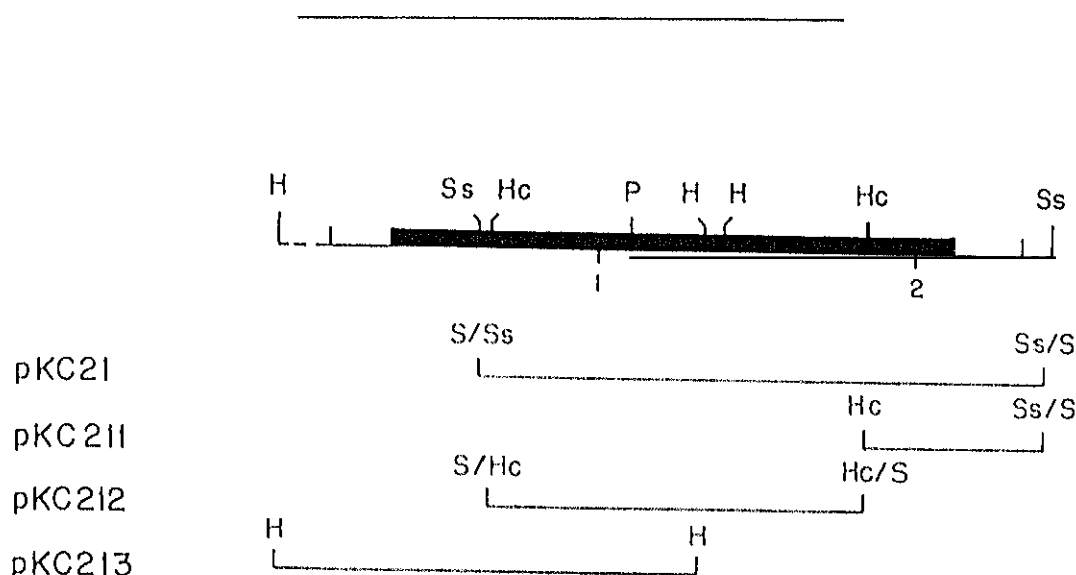


Figura 22- Mapa físico del gen *cspD*. En la parte superior de la figura se muestra el mapa físico del fragmento de DNA que contienen la secuencia completa del gen *cspD* resultado de la unión de los insertos de los plásmidos pKC21 y pKC213. Los trazos finos de la parte inferior de la figura hacen referencia a los tamaños de los insertos clonados y utilizados en la secuenciación del gen *cspC* y regiones no codificantes. Dianas de restricción: P, *PstI*; Ss, *SspI*; S, *SmaI*; H, *HindIII*; Hc, *HincII*.

La región 5' terminal del gen se aisló mediante una genoteca *HindIII* construida también en TG-1, siendo el vector de clonación un pUC18 digerido con *HindIII* y utilizando como sonda el fragmento *KpnI-PstI* de 459 pb del plásmido pKC21 (Fig. 22), el cual ya se había utilizado previamente para la realización de un *Southern-blot* que nos permitió seleccionar la enzima de restricción más adecuada para conseguir la clonación del gen completo (Fig. 23). Se obtuvo de esta manera un clon positivo con el plásmido recombinante pKC213 (Fig. 22).

Los fragmentos que codifican para el gen *cspD* se secuenciaron mediante la preparación de subclones de dichos plásmidos (Fig. 22) usando oligonucleótidos sintéticos y *primers* universales, dando como resultado la secuencia que se muestra en la figura 24. El

análisis de la secuencia de nucleótidos reveló que el gen está constituido por 1.752 pb que codifican para una proteína de 583 aminoácidos, correspondiendo los 26 primeros residuos al péptido señal. El M_r de la proteína sin procesar es de 62.877 y para la proteína madura de 62.217. A 9 pb del codon de iniciación ATG se encuentra la secuencia de un posible RBS. Posee también la proteína un dominio de unión a colina constituido por 5 motivos repetidos (Fig. 57, pag. 162) y un cola terminal de 8 aminoácidos (Fig. 24).

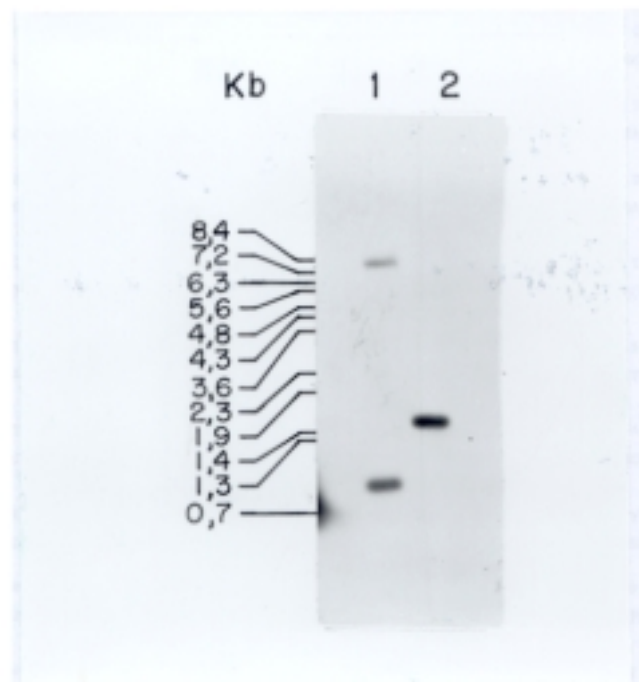


Figura 23- Análisis mediante *Southern-blot* del DNA de *C. acetobutylicum* NCIB 8052. El DNA se digirió con las enzimas *Pst*I (calle 1) y *Hind*III (calle 2). En el margen izquierdo de la figura se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (*λ*BstEI). Se utilizó como sonda el fragmento *Kpn*I-*Pst*I del plásmido pKC21. La temperatura de hibridación y lavado fue de 65°C.

2.4- Comparación de las secuencias de nucleótidos y aminoácidos de los genes *cspA*, *cspB*, *cspC* y *cspD*.

La comparación de las secuencias de nucleótidos de los genes *cspA*, *cspB*, *cspC* y *cspD* se muestra en la figura 25. Dicha comparación se ha llevado a cabo teniendo en cuenta el alineamiento en la comparación de aminoácidos que se muestra en la misma figura.

1 TAAATTCACCAAGTACTAATATTTTATTAATAAATATGACAAAAACCTCCAAATAAATACAAAAGACGTAGACAAGCA
RBS
83 TAAATAAAGTTGCTATAATAATATTCGTAACCCAGATAATATTTTAAACCAAGATATTATTTTAACTAATAAATAATTAATAAAGCAAAAATGAAACTCTAGGAGGAATATACAA
1' H I K R H T K A T S L L V A A A A I I S I V P A H A A D Y T K I D S Q E G T I Y
203 ATGATAAAAAGAAATGACTAAAGCAACATCGTTACTAGTAGCTGCAGCTGCTATTATTTCAATAGTACCTGCACATGACCTGACTACACAAAAATCGATTCAAGAAAGGTACTATCTAT
41' D A V A Y K D G K F Y V D G E V N D K D E A A Y Y L A D G K Y N N L S D I D S G
323 GATGCAGTAGCATATAAAGATGGTAAGTTCTATGTTGATGGAGAAGTTAATGACAAAGACGAAGCAGCTTATTATTTAGCAGATGGTAAATACAATAATTTATCAGATATTGATAGTGG
81' A D T D I Y G S K Y L E V Q D G D Y F V D L D N G S V T D E S I K E N A E D D A
443 GCAGATACTGATATTATGGATCTAAATATTTAGAAAGTACAAGATGGTACTACTTTGTTGACTTAGATAATGGTTCACTAAGTATGATGAATCTATAAAGAAAAATGCAGAAAGATGATGCA
121' A S A L R K N L K K D N D K R Y A S S E R E T I Q D L A G A E I A G N K F S A P
563 GCATCAGCTTTAAGAAAGAACCTTAAGAAAGATAATGATAAAGATATGCTTCAAGCGAACCTGAACAAATCAAGATTAGCTGGCGCAGAAATTCAGGAAATAAATCTCAGCTGCT
161' W Y K F E Y T K D N A T N G S V A K L N V Y T D A Q G N Y I D A D Y N L G S I K
683 TGGTATAAATTTGAATATACAAAAGATAATGCAACAAACGGTTGAGTTGCTAAATTAATGTTTATCTGATGCTCAAGGAAATTACATTGATGCTGATTACAATTAGGATCAATTA
201' V T T T A S S T T S K T V T I A N T D D S Y D A D G T S D A T G K K S S A S V D
803 GTAACACAACTGCATCTTCAACACTCTTAAACAGTAAGTATAGCTAATACAGATGATCTTATGATGCTGATGGTACTTCTGATGCTACTGGAAAGAAAGTTCAGCAAGTGTGTAT
241' T S A G K V I G Q D A N Y I Y R T A K V T V N A A H G A T I T K I N G H D V T V
923 ACTTCTGCAGGAAAGTAATTTGGACAAGATGCCAACTACATATACAGAACAGCAAGGTTACTGTAAATGCTGCTAATGAGCTACAATACTAAAAATAATGGAATGGATGTTACTGTT
281' D G G H T F T V S N G T L G S S A S T V G F N V I Q K I S K A Q A S G N V D G A
1043 GATGGTGGACATACATTACTGTAAACCAATGGTACATTAGGATCAAGTGTAGTACAGTTGGATTCAATGTTATCTCAAAAAATATCAAAAGCAAGCTTCAGGTAATGTTGATGGAGCT
321' F Y A K S V T T Y I V S D E D G K H E A F A Y D H Y T V A N G N I V G Y T A N G
1163 AAATATGCTAAATCAGTAACAACCTTATATAGTTTCTGATGAAGATGGAAAAATGAAGCTTTCCTTATGATACTATCTGTTGCTAACGAAAAATAGTTGGTTATCTGCTAATGGT
361' T N V N T A T G T L S B K N G Y Y Y V D L G D E A S E D V K E N G T K S A V D T
1283 ACAAATGTAATACTGCTACAGGAACATTAAAGCTCAAGAACGGTTATTACTATGTTGATTAGCGGATGAAGCAAGTGAAGATGTAAGAAAAATGGTACTAAGAGTCTGTTGATACA
401' D V D G K V W R L D A Q Y I Y K W D N D E D W T K V Y K V D G S F D Q H S V Y N
1403 GATGTAGATGGAAAGGTTTGGAGATTAGATGCTGGATACATCTACAAATGGGATAATGATGAAGATTGGACAAAAGTATACAAAGTAGATGGATCATTGATCAAAATGTCAGTATACAAC
441' Q D N I V A W S K E D D V Y S V I G G K G T S T P T D P T F V V N K G W V K T D
1523 CAAGATAATATCGTAGCTTGGAGTAAGAAGATGATGTTACTCAGTAATTTGGTGAAGGTACAGTACACCAACATATCTTACTCCAGTTGTAATAAAGGATGGGTTAAGACTGAT
481' A G H T F Y N T D P T Q V K G Q W V N D G G V W Y Y I K A D G T H A T G N I Q D
1643 GCAGGTTGGACATTCTACAATACTGATCCAACTCAAGTAAAGGACAATGGGTTAACGATGGTGGAGTTTGGTATTACATCAAGCTGATGGAATATGGCAACTGGTTGGATTCAAGAT
521' G S T W Y Y L Q S S G A H K T G W L N D N G T H Y Y L Q S S G A H S T G W L N D
1763 GGATCACTTGGTACTATTACAATCATCAGGACCAATGAAACTGGTTGGTTAAATGATAACGGAACCTGGTACTACTTACAATCATCAGGACCAATGTCTACTGGTTGGTTAAATGAT
561' N G T W Y F L N S S G A H L A N T T V D G Y N
1883 AATGGAACTTGGTATTCTTAAATTCATCAGGACCAATGTTAGCTAACCAACTGTTGATGGATACAATTAGGACATTAAGGCTTGGATTAAAGTAATAAATTAATTAATAGAGTT
2003 TAGATAATAAATTTAACTTCTACAACATAAATACTAAAGATAAAGATTACTTTTGAATAATTTCTATTGAAAGTGAATATAGGTAATGAAGGCTGCTCAATAATGGTTATTATTAATC
2123 GTTATGAGCAGTTTTTTTGGGTATCATTAAATAGTATTTTACAAAGTGTTAAATAA

Figura 24- Secuencia de nucleótidos y aminoácidos del gen *csuD*. Se muestra la secuencia de aminoácidos en código de una letra por encima de la secuencia de nucleótidos. El sitio de unión al ribosoma (RBS) y el posible terminador de transcripción aparecen subrayados. P1, P2, P3, P4, y P5 hacen referencia a los motivos de unión a colina presentes en la proteína CspD y T a un posible motivo imperfecto. El codón de terminación (TAG) se indica con un asterisco.

Cuando se cuantifica la semejanza entre las regiones estructurales de los 4 genes (Tabla 6) se observa que son muy similares. El porcentaje de identidad es muy parecido entre ellos siendo el máximo de un 65,6% entre los genes *cspB* y *cspC* y un mínimo de un 54,6% entre los genes *cspC* y *cspD*. En el extremo 5' de los genes estructurales, el mayor porcentaje de identidad es de un 67,7% entre los genes *cspA* y *cspB* y el mínimo de un 52% entre los genes *cspC* y *cspD*. Respecto al extremo 3' de los genes estructurales, el mayor porcentaje es de un 66,3% entre los genes *cspC* y *cspD* y el mínimo de un 52,5% entre los genes *cspC* y *cspA*. Aunque las regiones no codificantes de estos genes no manifiestan ese grado de similitud hay que destacar que los 43 primeros nucleótidos de la región 3' no codificante de los genes *cspA* y *cspC* son casi idénticos.

La estructura general de las cuatro proteínas es la misma, estando estructuradas en un dominio N-terminal y un dominio C-terminal, siendo éste último el dominio responsable de la unión a colina. El porcentaje de similitud e identidad a nivel de aminoácidos es muy elevado (Tablas 7 y 8). El máximo de similitud detectado es de un 80,2% entre las proteínas CspB y CspC, siendo el máximo de identidad de un 70,1% entre las proteínas CspA y CspC. El mínimo de similitud y de identidad es de un 68,1% y 54,8% respectivamente entre las proteínas CspC y CspD, al igual que ocurría a nivel de nucleótidos. En cuanto a los porcentajes de similitud e identidad entre los dominios N-terminales por un lado y los dominios C-terminales por otro se muestran también en la tablas 6 y 7, observándose que los máximos valores se alcanzan en la región N-terminal. El mayor porcentaje de similitud en el dominio N-terminal se consigue entre las proteínas CspC y CspB siendo de un 83,4%, mientras que a nivel de identidad, el mayor porcentaje se alcanza entre las proteínas CspA y CspC siendo de un 72,3%. Los mínimos tanto a nivel de similitud como de identidad son de un 67,5% y 52,5% entre las proteínas CspC y CspD, respectivamente. Respecto al dominio C-terminal, el máximo valor de similitud es de un 75% entre las proteínas CspA y CspD, siendo el mayor porcentaje de identidad de un 64,2% entre las proteínas CspC y CspD al contrario de lo que ocurría con estas proteínas a nivel del dominio N-terminal donde se alcanzaban los valores mínimos. Los mínimos son de un 66,9% a nivel de similitud entre los genes CspB y CspD y de un 57,6% de identidad entre los genes CspB y CspC. El dominio C-terminal está estructurado en motivos repetidos de 20 aminoácidos cada uno, variando el número de unas proteínas a otras, siendo de 4 motivos para la CspA, 5 motivos para la CspC y CspD y 6 motivos para la CspB, terminando todos ellos en una cola con un número

de aminoácidos variable a excepción del gen *cspB*. Hay que destacar que los motivos que se localizan en la misma posición en las cuatro proteínas Csp son más parecidos entre ellos que si se comparan los motivos presentes en cada proteína.

```

CspA M F K R A N K I T S L L V A A A S V M A L V P - - 23'
CspB M I R G M G K V T S L L V A A A T V A S L V P F S 25'
CspC - - - - - - - - - - - - - - - - - - - - - -
CspD M I K R M T K A T S L L V A A A A I I S I V P - - 23'
      *               * * * * * * * * * * * * * *

```

```

ATGTTTAAAAGAGCAACAAAATTACATCTTTATTAGTAGCTGCTGCTTCAGTTATGGCTTTAGTGCCT----- 392
ATGATAAGAGGCATGGGAAAGGTAACATCACTATTAGTAGCAGCAGCAACAGTTGCTTCATTAGTACCTTTTAGT 240
-----
ATGATAAAAAGAATGACTAAAGCAACATCGTTACTAGTAGCTGCAGCTGCTATTATTTCAATAGTACCT----- 271
*** * * * * ** * * * * * * * * * * * * * * * * * * * * *

```

```

CspA - A Y A A D V K K V D S E D G T V Y N A V A Y K D 47'
CspB G A N A A E V K R I S S D D G T I Y N A I A Y K D 50'
CspC - - - - A D V K K I D S D D G T V Y N A V A Y K D 21'
CspD - A H A A D Y T K I D S Q E G T I Y D A V A Y K D 47'
      * * * * * * * * * * * * * * * * * * * * *

```

```

---GCTTATGCTGCAGACGTAAGAAAGTTGATTTCAGAAGATGGTACTGTATACAATGCAGTAGCATACAAGAT 464
GGCGCTAATGCTGCTGAGGTGAAAAGAATTAGTTCTGATGATGGAACATCTATAATGCAATAGCATATAAGAC 315
-----GCAGATGTAAAAAAATTGATTTCAGATGATGGTACTGTATACAATGCAGTAGCTTACAAGAT 65
---GCACATGCAGCTGACTACACAAAAATCGATTTCACAAGAAGGTACTATCTATGATGCAGTAGCATATAAGAT 343
** * * * * * * * * * * * * * * * * * * * * * * * * * *

```

```

CspA G K Y F V D G E I N D - D E E A Y Y V A D G K F N 71'
CspB G K A Y I D G E I N D - D E E A Y Y L S N G K F N 74'
CspC G K F F V D G E I N D - D E E A Y Y V S D G K F N 45'
CspD G K F Y V D G E V N D K D E A A Y Y L A D G K Y N 72'
      * * * * * * * * * * * * * * * * * * * * *

```

```

GGTAAATACTTTGTTGATGGAGAAATCAACGAT---GATGAAGAAGCTTACTATGTAGCTGACGGAAAATTCAAC 536
GGTAAAGCTTATATCGATGGAGAAATTAATGAT---GACGAAGAAGCTTACTATTTATCCAACGGAAAATTCAAT 387
GGTAAATTCTTTGTTGATGGAGAAATTAACGAT---GACGAAGAAGCCTACTATGTATCTGATGGAAAATTTAAC 137
GGTAAGTTCTATGTTGATGGAGAAGTTAATGACAAAGACGAAGCAGCTTATTATTTAGCAGATGGTAAATACAAT 418
***** * * * * * * * * * * * * * * * * * * * * * * *

```

```

CspA K L E D V D S G D E A V L F G E K Y L D V S D G D 96'
CspB K L D D V D S G D S A A L F G E K Y L D I S D G D 99'
CspC K L E D I D S G D D A D L F G E K Y L D I N D G D 70'
CspD N L S D I D S G A D T D I Y G S K Y L E V Q D G D 97'
      * * * * * * * * * * * * * * * * * * * * *

```

```

AAATTAGAAGATGTTGATTTCAGGAGATGAAGCAGTTCTATTGCGAGAAAAATACTTAGATGTATCAGACGGAGAC 611
AAATTAGACGATGTTGATTTCAGGAGATAGCGCAGCTCTATTGCGAGAAAAATATTTAGATATATCAGATGGAGAC 462
AAACTAGAAGATATTGATTTCAGGAGATGACGCTGATCTATTGCGAGAAAAATATCTTGATATAAATGATGGTGAT 212
AATTATCAGATATTGATAGTGGAGCAGATACTGATATTTATGGATCTAAATATTTAGAAGTACAAGATGGTGAC 493
***** * * * * * * * * * * * * * * * * * * * * * * *

```

CspA Y T V D L D K G S V T D D D V K G D T A D D A A A 121'
 CspB Y T V D L D K G S V T D D D I K G D T E D D A A A 124'
 CspC Y T V D L D K G S I T D D D T K G D T A D D A S A 95'
 CspD Y F V D L D N G S V T D E S I K E N A E D D A A S 122'

* * * * *
 TACTACTGTTGATTTAGATAAAAGGTAGTGTAACTGATGATGACGTTAAAGGTGATACAGCAGACGATGCTGCAGCA 686
 TATACTGTTGACTTTAGATAAAAGGTAGCGTAACTGATGATGATATTAAAGGTGATACTGAAGATGATGCCGCTGCT 537
 TACTACTGTTGATTTAGATAAAAGGTAGTATAACTGATGATGATACTAAGGGCGATACTGCTGATGACGCTTCAGCA 287
 TACTTTGTTGACTTTAGATAATGGTTCAGTAACTGATGAATCTATAAAAGAAAATGCAGAAGATGATGCAGCATCA 568

CspA A L R K K I K D D T D D R Y L E S E A E G V K N E 146'
 CspB A L R K K I K D D T D D R Y N E T E A N T I K D S 149'
 CspC D L R K K I K D D T D D R Y D E D F A D N I I D E 120'
 CspD A L R K N L K K D N D K R Y A S S E R E T I Q D L 147'

* * * * *
 GCTTTAAGAAAGAAAATTAAAGATGATACAGATGATAGATATTTAGAAAGTGAAGCAGAAGGCGTTAAAAATGAA 761
 GCTTTAAGAAAAAAATCAAAGATGATACAGATGATAGGTATAATGAACTGAAGCAAATACTATTAAAGATTCA 612
 GATCTAAGAAAGAAAATTAAAGATGACACTGATGACAGATATGATGAAGATTTTGCTGATAATATAATAGATGAA 362
 GCTTTAAGAAAGAACCTTAAGAAAGATAATGATAAAAGATATGCTTCAAGCGAACGTGAACAATTCAAGATTTA 643

CspA D D - - - - - L A I I G G A K Y D K P W Y A 163'
 CspB N H G D - - - - - L F D L I P G A K Y N K V W Y Y 169'
 CspC N K D D T E D L T K L N I I P G A K Y K T P W Y Y 145'
 CspD A G A E - - - - - I A G N K F S A P W Y K 163'

* * * * *
 GATGAT-----CTTGCTATAATTGGAGGAGCAAAGTATGACAAGCCTTGCTATGCT 812
 AATCATGGTGAC-----CTATTGCTTTAATTCAGGAGCAAAGTATAATAAGGTTTGGTATTAT 672
 AATAAAGATGATACTGAAGATTTAACTAAATTAAATATAATTCAGGTGCTAAATATAAACTCCATGGTATTAT 437
 GCTGGCGCAGAA-----ATTGCAAGGAAATAAATTCTCAGCTCCTTGCTATAAA 691
 * * * * *

CspA T T Y T A S A K A I G D - V N G L T A T N N K - - 185'
 CspB T Q Y K A A Q K S I D K N V N G L N G L D - S A H 193'
 CspC T Q Y K A A D K A V N D N V N G L K E I T D Q N K 170'
 CspD F E Y T - K D N A T N G S V A K L - - - - - 179'

* * * * *
 ACAACATATACTGCTTCTGCAAAAGCTATAGGTGAT---GTTAATGGATTAAACAGCAACAAATAATAAG----- 878
 ACACAAATATAAAGCCGCTCAAAAATCTATAGATAAAAATGTCAATGGATTAAATGGATTGGAT---TCAGCCCAT 744
 ACACAAATATAAAGCAGCAGATAAAGCTGTGAATGATAATGTTAATGGACTAAAGGAAATAACTGATCAAAATATAA 512
 TTTGAATATACA---AAAGATAATGCAACAAACGGTTTCAGTTGCTAAATTA----- 739

CspA - - F N V Y T D T N G A Y I D A D Y N L G K V K V 208'
 CspB Q I F N V F T D D K G N Y I D A D Y N L G K V K V 218'
 CspC R V F N V Y T D T K G G Y I D A D Y N L G K V K V 195'
 CspD - - - N V Y T D A Q G N Y I D A D Y N L G S I K V 201'

* * * * *
 -----TTCAATGTTTATACTGATACAAATGGTGCTTACATTGATGCAGATTATAACTTAGGAAAAGTAAAGTT 947
 CAAATATTTAATGTATTACTGATGATAAAGGCAATTATATTGATGCGGACTATAACTTAGGTAAAGTTAAGGTT 819
 CGAGTTTTCATGTTTATACTGACACAAAAGGTGGCTACATTGATGCTGACTATAACTTAGGAAAAGTTAAGGTT 587
 -----AATGTTTATACTGATGCTCAAGGAAATTACATTGATGCTGATTACAATTTAGGATCAATTAAAGTA 805
 **

CspA T T T A D G A - - - - T K E V T V E N T N D T Y 228'
 CspB T T T A S S A S G T T L T K T D T I E N T N D A Y 243'
 CspC K T T A D G A S - - - - T K T E T I E N T N D K Y 216'
 CspD T T T A S S T T - - - - S K T V T I A N T D D S Y 222'
 * * * * *

ACAACAACAGCTTCAAGTGCTTCTGGTACTACATTAACAAAGACAGATACAATAGAGAACACAAACGACGCTTAC 1007
 ACAACAACAGCTGACGGTGCT-----ACAAAAGAAGTTACTGTAGAAAACACAAACGATACTTAT 894
 AAAACCACTGCTGATGGTGCATCA-----ACTAAACAGAAACAATAGAAAATACTAATGATAAATAC 650
 ACTACAACAGCTGATCTTCAACAAC-----TCTAAACAGTAACATATAGCTAATACAGATGATTCTTAT 868

CspA D A A G S - - - - S D S G K D K V S A S V - - K 246'
 CspB D A A D G I I N G T S V S G S D K L S A S V - - V 266'
 CspC D A A G - - - - - - - - K D D A L T A S V - - S 230'
 CspD D A D G - - - - T S D A T G K K S S A S V D T S 242'
 * * * * *

GATGCAGCTGGTTCT-----AGCGATTGAGGAAAGATAAGTAAGTGCAAGTGTT-----AAA 1061
 GATGCAGCTGATGGAATTATCAATGGAAGTAGCGTTAGCGGCTCTGATAAATTAAGTGCAAGTGTT-----GTT 963
 GATGCTGCTGGT-----AAGGATGATGCTAAGTGCAAGTGTT-----TCC 692
 GATGCTGATGGT-----ACTTCTGATGCTACTGGAAGAAAGTTGCAAGTGTTGATACTTCT 928

CspA Q T A V L T Q D K D N I Y R L - - V E V T V K T G 269'
 CspB Q D R V L T Q D K D Y I Y R L A T V K I T I T T G 291'
 CspC Q D K I L T Q D D D Y I Y R L A T I K V N V T T D 255'
 CspD A G K V I G Q D A N Y I Y R T A K V T V N A A N G 267'
 * * * * *

CAAACCTGCAGTTTAAACACAAGATAAGGATAATATCTACAGACTT-----GTAGAAGTAAGTGTTAAACAGGT 1130
 CAAGATAGAGTATTAACCCAAAGATAAGATTATATATATAGACTTGCAACTGTAAATAAATATTACCACCTGGT 1038
 CAAGATAAATATTGACACAAGATGATGATTATATTTATAGACTTGCGACTATAAAGTAATGTAAGTACTGAT 767
 GCAGGAAAGTAATGGAACAAGATGCAAACTACATATACAGAACAGCAAGGTTACTGTAAATGCTGCTAATGGA 1003

CspA H G V - - - I T E I N G V K I A D I G A N T - - - 288'
 CspB A A A T - - I S E I N G V K V D P N N S N D - - - 311'
 CspC A G A K V R I S E I N G V E L N P K N A N D K P Q 280'
 CspD A T - - - I T K I N G M D V T V D G G H T - - - 285'
 * * * * *

CATGGTGTT-----ATAACAGAAATTAATGGCGTAAAAATAGCTGACATTGGTGCTAATACA----- 1187
 GCAGCTGCTACA-----ATTAGTGAAATAAATGGTGTTAAAGTTGATCCAAATAATAGCAACGAT----- 1098
 GCTGGTGCTAAAGTTAGAATAAGTGAAATAAATGGTGTTGAACCTTAACCCAAAAACGCAATGACAAACCACAA 842
 GCTACA-----ATAACTAAAATAAATGGAATGGATGTTACTGTTGATGGTGACATACA----- 1057

CspA V F G G T - - - - N - - T E V T F P A I Q K I S 306'
 CspB I F K V E N N - - - - G Q V V S F K A I Q K I S 331'
 CspC I F K V E N D - - - - G G T V S F K A I Q K I S 300'
 CspD - F T V S N G T L G S S A S T V G F N V I Q K I S 309'
 * * * * *

GTATTTGGTGGTACA-----AAC-----ACTGAAGTAACATTCACAGCTATCCAAAGATATCA 1241
 ATCTTCAAAGTTGAAAATAAC-----GGACAAGTTGTATCTTCAAAGCTATTCAAAAAATATCA 1158
 ATATTTAAAGTTGAAAATGAT-----GGAGGAAGTGTATCATTCAGGCTATTCAAAAGATATCC 902
 ---TTTACTGTAAGCAATGGTACATTAGGATCAAGTGCTAGTACAGTTGGATTCAATGTTATCCAAAAATATCA 1129

CspA K A Q A S D D V D G A K Y A K T V T T Y A L S D D 331'
 CspB K T Q A S G D I D D A K Y A K T V T T Y A L S D K 356'
 CspC K A Q D S D D V D G A K Y A K T V T T Y A L S D E 325'
 CspD K A Q A S G N V D G A K Y A K S V T T Y I V S D E 334'

* * * * *
 AAAGCACAAGCTTCTGATGATGTTGATGGAGCTAAATATGCTAAAAACAGTAACACTATTATGCTTTATCTGATGAT 1316
 AAAACTCAAGCTTCTGGTGATATTGATGACGCTAAATATGCTAAAACTGTAACACTATTATGCACCTTCTGATAAA 1233
 AAAGCTCAAGACTCCGATGATGTTGATGGAGCTAAATATGCAAAAACTGTAACACTATTATGCACCTTCTGATGAA 977
 AAAGCACAAGCTTCAGGTAATGTTGATGGAGCTAAATATGCTAAATCAGTAACAACCTTATATAGTTTCTGATGAA 1204

CspA S G N K L D E Q S L F V K T D G T A V T T T K Y T 356'
 CspB D G K K L D A E D L F I N T S G N V V T T T N Y T 381'
 CspC D G K N L D A K D L F L N T D G T I N E D T K F T 350'
 CspD D G K N - - - - - E A F A Y D N Y T 347'

* * * * *
 TCAGGAAATAAATTAGATGAACAAAGTTTATTTGTAAAACTGATGGTACTGCTGTTACTACAACCTAAGTATACT 1391
 GATGGTAAAAAGTTAGATGCTGAAGATTTATTTATAAATACTTCAGGCAACGTAGTCACTACAACCTAAGTATACT 1308
 GATGGTAAAAACCTAGATGCTAAAGACTTATTCCTAAATACTGATGGCACTATCAATGAAGATACTAAATTTACT 1052
 GATGGAAAAAAT-----GAAGCTTTCGCTTATGATAACTATACT 1243

CspA V V N G K L I A Y N T D I N D N K K V T V D A Y T 381'
 CspB V G S G K L I A Y N S E I N N N D K V T V R A Y T 406'
 CspC V A N G K L I A Y N T E M N E Y K K V V V R A Y S 375'
 CspD V A N G K I V G Y T A N G T N V N T A T G - - - T 369'

* * * * *
 GTAGTCAATGGAAAGCTTATAGCTTACAATACTGATATTAATGACAATAAAAAAGTAACTGTTGATGCATATACA 1566
 GTAGGTTCTGGAAAGCTTATAGCTTACAATTCGAAATTAACAATAACGACAAAGTTACTGTTAGAGCTTATACA 1383
 GTAGCTAACGGAAAACTTATAGCCTACAATACAGAAATGAATGAGTATAAAAAAGTTGTTGTGAGAGCATATTCA 1127
 GTTGCTAACGGAAAAATAGTTGGTTATACTGCTAATGGTACAAATGTAAATACTGCTACAGGA-----ACA 1309

CspA L K T K G G Y Y Y A D E E D K S E E D C E V S A Q 406'
 CspB L K S S S G F Y Y A D E E D Q S K E D C E N S K N 431'
 CspC L K T K G G Y Y Y A D E E D E S K E D C E V S S Q 400'
 CspD L S S K N G Y Y Y V D L G D E A S E D V - - - K E 391'

* * * * *
 TTAAAGACAAAAGGTGGATACTACTACGCTGATGAAGAAGATAAGAGCGAAGAAGATTGCGAAGTAAGTGCTCAA 13341
 TTAAATCAAGTAGTGGATTCTATTATGCTGATGAGGAAGATCAAAGCAAAGAAGATTGTGAAAACAGCAAAAAC 1458
 TTAAAAACAAAAGGTGGATACTACTACGCCGACGAAGAAGATGAAAGTAAAGAAGATTGTGAAGTAAGTTCTCAA 1202
 TTAAGCTCAAAGAACGGTTATTACTATGTTGATTAGGCGATGAAGCAAGTGAAGATGTA-----AAAGAA 1375

CspA D K T T A A V Q T D V D G N L W R L D G G Y I Y K 431'
 CspB Q G - - A A V Q T D V D G N L W R L D G G Y I Y K 454'
 CspC D D K T S A V Q T D V D G N L W R L D G G Y I Y K 425'
 CspD N G T K S A V D T D V D G K V W R L D A G Y I Y K 416'

* * * * *
 GATAAAACAACAGCAGCTGTTCAAACAGACGTTGATGGAAATCTTTGGAGATTAGATGGTGGATATATCTACAAA 1616
 CAAGGT-----GCTGCTGTTCAAACAGATGTTGACGGAAATCTATGGAGATTAGATGGTGGATATATCTACAAA 1527
 GATGATAAGACATCAGCTGTTCAAACAGACGTTGATGGAAACCTATGGAGATTAGATGGTGGATATATCTATAAA 1277
 AATGGTACTAAGAGTGCTGTTGATACAGATGTAGATGGAAAGGTTTGGAGATTAGATGCTGGATACATCTACAAA 1450
 *** **

CspA F D N T D D W D K V Y K V D G S F D E L S V Y D K 456'
 CspB F D N T D D W D K V Y R V D G S F D E F S V Y D K 479'
 CspC F D N T D D W D K V Y K V D G S F D E L S V Y D K 450'
 CspD W D N D E D W T K V Y K V D G S F D Q M S V Y N Q 441'
 * * * * *

TTTCGATAACACTGATGATTGGGATAAAGTTTACAAAGTAGACGGATCTTTTCGATGAATTATCAGTTTACGATAAA 1691
 TTTTGATAAATCTGATGACTGGGACAAAGTATACAGAGTAGACGGATCATTTGATGAATTCTCAGTTTATGATAAA 1602
 TTTTGATAATACAGATGATTGGGACAAAGTTTACAAAGTCGATGGATCTTTTGATGAGCTTTCAGTCTATGATAAA 1352
 TGGGATAATGATGAAGATTGGACAAAAGTATACAAAGTAGATGGATCATTTGATCAAATGTCAGTATACAACCAA 1525

CspA D N M V A W S E D D D V Y S L I G G K S D D N K G 481'
 CspB D N I V A W S K D D D V Y S L I G G K Q S N T D P 504'
 CspC D N I A A W S E D D D V Y S L I G G K S S D D N N 475'
 CspD D N I V A W S K E D D V Y S V I G G K G T S T P T 466'
 * * * * *

GACAACATGGTTGCTTGGAGCGAAGATGATGATGTATATTCATTAATCGGAGGAAAGAGCGATGATAATAAGGGT 1766
 GATAATATCGTTGCTTGGAGCAAAGATGATGATGTTTATTCATTAATAGGTGGAAAACAATCTAATACTGACCT 1677
 GACAATATAGCTGCTTGGAGTGAAGATGATGATGTTTATTCATTAATAGGAGGAAAACTTCTGATGATAATAAC 1427
 GATAATATCGTAGCTTGGAGTAAAGAAGATGATGTTTACTCAGTAATTGGTGGTAAAGGTACAAGTACACCAACA 1600
 ** *****

CspA D D Q G T T P V V K A G W A Q T S A G - W T Y V K 505'
 CspB D D - - - T P V V K T G W V Q A S D G T W S Y I K 526'
 CspC T N - - - P P T T T A G W V Q A A D G T W T Y N K 497'
 CspD D - - - P T P V V N K G W V K T D A G - W T F Y N 487'
 * * * * *

GATGACCAAGGAACAACCTCCTGTAGTTAAAGCTGGTTGGGCTCAAACCTTCAGCAGGA---TGGACTTATGTTAAA 1838
 GATGAT-----ACTCCAGTAGTAAAAACAGGATGGGTACAAGCTTCAGATGGAACCTTGGTCTTATATATAAA 1743
 ACTAAT-----CCTCCAACCTACAACCTGCAGGATGGGTACAAGCTGCTGATGGAACCTTGGACTTATAATAAA 1493
 GAT-----CCTACTCCAGTTGTAATAAAGGATGGGTTAAGACTGATGCAGGT---TGGACATTCTACAAT 1663
 *** *****

CspA A D G T K A T G - W L Q D G G A W Y Y L K A D G T 529'
 CspB A D G N K A T G - W V Q D G S T W Y Y F K A D G S 550'
 CspC E D G T K A T G - W L N L N G V W Y Y L N T D G V 521'
 CspD T D P T Q V K G Q W V N D G G V W Y Y I K A D G T 512'
 * * * * *

GCTGATGGAACATAAGCTACTGGA---TGGTTACAAGACGGTGGTGGTACTACTTAAAGCTGATGGGTACA 1910
 GCTGACGGAATAAGGCGACAGGT---TGGGTTCAAGATGGATCAACTTGGTACTATTTTAAAGCTGATGGATCT 1815
 GAAGATGGAACATAAGGCTACTGGC---TGGTTAAACTTAAATGGAGTTTGGTATTATTTAAACACTGATGGTGT 1565
 ACTGATCCAACCTCAAGTAAAAGGACAATGGGTTAACGATGGTGGAGTTTGGTATTACATCAAAGCTGATGGAAC 1738

CspA M A T G W I Q D G A T W Y Y L N G S G A M Q T G W 554'
 CspB M A T G W V Q D G S T W Y Y F Q S W G G M Q T G W 575'
 CspC M A T G W L N L N G V W Y Y L N P S G A M A T S W 546'
 CspD M A T G W I Q D G S T W Y Y L Q S S G A M K T G W 537'
 * * * * *

ATGGCTACAGGTTGGATTCAAGATGGAGCAACTTGGTACTACTTAAACGGATCAGGTGCTATGCAAACCTGGTTGG 1985
 ATGGCTACTGGTTGGGTTCAAGATGGATCAACTTGGTATTACTTCCAATCATGGGGTGAATGCAAACCTGGTTGG 1890
 ATGGCTACTGGTTGGTTAAACTTAAATGGAGTTTGGTATTACTTAAATCCATCAGGAGCTATGGCTACTAGCTGG 1640
 ATGGCAACTGGTTGGATTCAAGATGGATCAACTTGGTACTATTTACAATCATCAGGAGCAATGAAAACCTGGTTGG 1813

```

CspA L N D N G T F Y Y L N G S G A M L S N - - - - - 573'
CspB L N N N G T W Y Y L Q S W G G M K T G W L N D N G 600'
CspC L N L N G T W Y Y L N S S G A M A T G W L N D N G 571'
CspD L N D N G T W Y Y L Q S S G A M S T G W L N D N G 562'
  * * * * *
TTAAATGATAACGGAACCTTCTACTACTTAAATGGATCAGGTGCTATGTTATCTAAC----- 2042
CTAAATAATAACGGTACTTGGTATTACTTACAATCATGGGGCGGAATGAAGACTGGTTGGCTAAATGATAACGGC 1965
TTAAACTTAAATGGAACCTTGGTACTATTAAATTCATCAGGAGCTATGGCTACTGGTTGGTTAAATGATAATGGA 1715
TTAAATGATAACGGAACCTTGGTACTACTTACAATCATCAGGAGCAATGTCTACTGGTTGGTTAAATGATAATGGA 1888
*****

CspA - - - - - T T T P D G Y Y V 582'
CspB N W Y Y L Q S W G G M Q T G W F N D N G T W Y Y L 625'
CspC T W Y Y L N S S G A M - - - L Y N T T V D G Y V L 593'
CspD T W Y F L N S S G A M - - - L A N T T V D G Y N - 583'
  * * * * *
-----ACAACAACCTCTGATGGATACTATGTA 2069
AATTGGTACTACTTACAATCATGGGGCGGAATGCAAACTGGTTGGTTTAAATGATAATGGTACTTGGTACTACCTA 2040
ACTTGGTATTATTAAATTCATCAGGGGCTATG-----TTATATAACACAACCTGTTGATGGTTATGTATTA 1781
ACTTGGTATTCTTAAATTCATCAGGAGCAATG-----TTAGCTAACACAACCTGTTGATGGATACAAT--- 1951
* * * * *

CspA G A N G A W V K - - 590'
CspB Y S N G V M A A N T 636'
CspC G S N G A W I R - - 601'
CspD - - - - -
  * * * * *
GGAGCTAATGGAGCTTGGGTAAA----- 2093
TACTCAAATGGTGTTATGGCCGCTAACACG 2073
GGCTCTAATGGAGCTTGGATTAGA----- 1805
*****

```

Figura 25- Comparación de las secuencias de aminoácidos de las proteínas Csp. Debajo de la secuencias de aminoácidos se muestran las secuencias de nucleótidos que codifican para las proteínas Csp. Cada asterisco significa la conservación de tres nucleótidos o tres aminoácidos en dicha posición. Los dominios de unión colina de las proteínas Csp se señalan en negrita.

Tabla 6- Porcentaje de identidad entre las regiones estructurales de los genes *cspA*, *cspB*, *cspC* y *cspD*.

Genes	<i>cspA</i>	<i>cspB</i>	<i>cspC</i>	<i>cspD</i>
A)				
<i>cspA</i>	--	65, 3	61	54, 6
<i>cspB</i>	65, 3	--	65, 6	58, 2
<i>cspC</i>	61	65, 7	--	54, 6
<i>cspD</i>	55, 6	58, 2	54, 6	--
B)				
<i>cspA</i>	--	67, 7	63	55, 5
<i>cspB</i>	67, 7	--	66, 3	57
<i>cspC</i>	63	66, 3	--	52
<i>cspD</i>	55, 5	57	52	--
C)				
<i>cspA</i>	--	54, 7	52, 5	56
<i>cspB</i>	54, 7	--	62, 6	63, 6
<i>cspC</i>	52, 5	62, 6	--	66, 3
<i>cspD</i>	56	63, 6	66, 3	--

A) Comparación de las regiones estructurales completas.

B) Comparación de las regiones estructurales que codifican para los dominios N-terminales.

C) Comparación de las regiones estructurales que codifican para los dominios C-terminales.

Tabla 7- Porcentaje de similitud entre las proteínas CspA, CspB, CspC y CspD..

Proteínas	CspA	CspB	CspC	CspD
A)				
CspA	--	79,1	77,6	70,5
CspB	79,1	--	80,2	71,1
CspC	77,6	80,2	--	68,1
CspD	70,5	71,1	68,1	--
B)				
CspA	--	80	78,8	70,4
CspB	80	--	83,4	72,1
CspC	78,8	83,4	--	67,5
CspD	70,4	72,1	67,5	--
C)				
CspA	--	74,4	74,2	75
CspB	74,4	--	67,7	66,9
CspC	74,2	67,7	--	70,6
CspD	75	66,9	70,6	--

A) Comparación de las proteínas completas.

B) Comparación de los dominios N-terminales de las proteínas.

C) Comparación de los dominios C-terminales de las proteínas.

Tabla 8- Porcentaje de identidad entre las proteínas CspA, CspB, CspC y CspD

Proteínas	CspA	CspB	CspC	CspD
A)				
CspA	--	68,2	70,1	56,8
CspB	68,2	--	68,7	55,3
CspC	70,1	68,7	--	54,8
CspD	56,8	55,3	54,8	--
B)				
CspA	--	70	72,3	55,9
CspB	70	--	71,4	54,6
CspC	72,3	71,4	--	52,5
CspD	55,9	54,6	52,5	--
C)				
CspA	--	59,1	61,8	61,3
CspB	59,1	--	57,6	58,7
CspC	61,8	57,6	--	64,2
CspD	61,3	58,7	64,2	--

A) Comparación de las proteínas completas.

B) Comparación de los dominios N-terminales de las proteínas.

C) Comparación de los dominios C-terminales de las proteínas.

5- Utilización de codones de los genes *cspA*, *cspB*, *cspC* y *cspD*.

Es bien conocido el bajo porcentaje en G+C que presenta el DNA de *C. acetobutylicum*. Este hecho ha ejercido un marcado efecto en la utilización de codones con A ó T, predominantemente (Young y cols., 1989). En la tabla 9 se exponen la utilización de codones de los genes descritos hasta el momento, no mostrándose ninguna diferencia significativa en cuanto a la utilización de codones entre unos genes y otros. Mayoritariamente se utilizan los codones con un alto contenido en A y T, siendo los más empleados los que presentan una A ó T en la tercera posición. Es de destacar la única utilización del codon AGA para la arginina, a excepción de los genes *cspB*, *cspC* y *cspD* que utilizan también, aunque una sola vez, los codones AGG, CGA y CGT, respectivamente. Respecto a la histidina y glutamina, sólo se utilizan los codones CAT y CAA, respectivamente y como codon de terminación los cuatro genes utilizan TAG. Al igual que ocurría con el gen *cspA*, las regiones codificantes son las que presentan un mayor contenido en G+C, siendo la media del 35% mientras que, en las regiones 5' y 3', la media es de un 19% y un 25%, respectivamente.

2.6- Estudio de la transcripción de los genes *cspA*, *cspB*, *cspC* y *cspD*.

Dada la presencia de cuatro genes homólogos, o isógenos, en la misma cepa bacteriana, podría ocurrir que dichos genes se transcribieran conjuntamente, en caso de que formaran parte de un operón, se transcribieran de manera independiente o no se transcribieran, al menos en las condiciones utilizadas. Para realizar el mapa de transcripción se llevó a cabo el aislamiento de RNA de *C. acetobutylicum* NCIB 8052 y posterior análisis del mismo por *Northern-blot* siguiendo las pautas descritas en el apartado 20.3 de Materiales y Métodos. Hay que destacar que las sondas utilizadas para realizar los experimentos de hibridación fueron los fragmentos *PvuII-SspI* del plásmido pKDE543, *HindIII* del pKC33, *PstI* del pKC31 y *PstI-EcoRI* del pKC21, pertenecientes a los genes *cspA*, *cspB*, *cspC* y *cspD*, respectivamente (Figuras 5,16,19 y 22). Como se puede observar en la figura 26, en las condiciones utilizadas, sólo se transcriben los genes *cspA* y *cspC* durante la fase exponencial de crecimiento, siendo el transcrito primario de un tamaño aproximado de 1,9 kb. Dicho tamaño se corresponde muy bien con el tamaño del transcrito primario del gen *cspA* de 1,8 kb, calculado en base al sitio de iniciación de la transcripción

Tabla 9- Utilización de codones de los genes *cspA*, *cspB*, *cspC* y *cspD*

AA	CODÓN	PROTEÍNA				AA	CODÓN	PROTEÍNA				AA	CODÓN	PROTEÍNA			
		CspA	CspB	CspC	CspD			CspA	CspB	CspC	CspD			CspA	CspB	CspC	CspD
F	TTT	5	9	7	4	S	TCT	7	12	7	9	Y	TAT	17	28	28	22
	TTC	6	7	5	6		TCC	0	1	3	0		TAC	22	13	13	19
L	TTA	24	20	21	19		TCA	12	20	14	26	X	TAA	0	0	0	0
	TTG	0	1	1	0		TCG	0	0	0	1		TAG	1	1	1	1
	CTT	4	3	6	1	P	CCT	4	2	1	3	H	CAT	1	2	0	2
	CTC	0	0	0	0		CCC	0	0	0	0		CAC	0	0	0	0
	CTA	1	7	8	1		CCA	1	3	6	3	N	AAC	12	15	14	10
	CTG	0	0	0	0		CCG	0	0	0	0		AAT	20	33	34	32
I	ATA	6	13	16	10	T	ACG	0	1	0	0	K	AAG	13	9	9	9
	ATC	5	7	2	6		ACA	30	18	12	25		AAA	40	41	46	35
	ATT	8	13	9	11		ACC	0	2	1	0	D	GAT	64	63	77	58
M	ATG	6	7	5	8		ACT	30	33	40	33		GAC	15	16	14	5
V	GTT	27	23	22	24	A	GCT	42	33	31	32	E	GAA	24	15	24	18
	GTC	1	2	3	0		GCC	0	4	3	0		GAG	0	3	2	0
	GTA	24	16	12	21		GCA	21	13	13	27	Q	CAA	12	20	10	14
	GTG	1	1	1	0		GCG	0	2	1	0		CAG	0	0	0	0
R	AGG	0	1	0	0	S	AGC	4	8	1	3	G	GGC	2	8	5	2
	AGA	5	8	7	5		AGT	6	8	6	9		GGT	25	23	17	22
	CGT	0	0	0	1	C	TGT	0	1	1	0		GGA	30	28	26	32
	CGC	0	0	0	0		TGC	1	0	0	0		GGG	0	0	1	0
	CGA	0	0	1	0	W	TGG	12	19	15	15						
	CGG	0	0	0	0	X	TGA	0	0	0	0						

X=codón de terminación

(Fig. 8) y al bucle de terminación (Fig. 7). La señal radiactiva es muy intensa para ambos genes, lo cual permite suponer que la cantidad de transcrito sintetizado es muy alta, lo que se podría traducir en la producción de una cantidad considerable de proteína. Esto en principio concuerda bien con el caso de la proteína CspA que como sabemos es la proteína mayoritaria del sobrenadante de cultivo de *C. acetobutylicum* NCIB 8052. En el caso del gen *cspB*, aparece una tenue banda del mismo tamaño que en casos anteriores, pero que pensamos que es inespecífica, pues los experimentos de *primer extension* realizados para localizar la zona promotora han sido infructuosos. Tampoco sería de extrañar que la hibridación fuera inespecífica dada la elevada homología que presentan estos genes a nivel de nucleótidos. Por otro lado, el hecho de que el gen *cspC* se transcribiera de manera importante nos hizo pensar en la posibilidad de que la segunda proteína mayoritaria presente en el medio de cultivo de *C. acetobutylicum*, con un M_r de 125.000, fuera la proteína CspC. Esta proteína se purifica también por cromatografía de afinidad en DEAE-celulosa junto con la proteína CspA, reacciona frente a los anticuerpos anti-CspA (Fig. 10) y da positivo el test de detección de glicoproteínas (Fig. 13). Para comprobar que nuestras conclusiones eran correctas, se purificó esta proteína por electroelución a partir de un extracto purificado en DEAE-celulosa y se determinó la secuencia de su extremo N- terminal, confirmandose este resultado.

2.7- Localización cromosómica de los genes *cspA*, *cspB*, *cspC* y *cspD*.

Del estudio de la transcripción de los cuatro genes se comprobó que dichos genes no constitúan un operón y que por tanto, existía la posibilidad de que ni siquiera se localizaran juntos en el mismo fragmento del cromosoma. Para ello se realizó una electroforesis en campo pulsado con el DNA cromosómico de *C. acetobutylicum* NCIB 8052. En la elección de las enzimas de restricción para la digestión del DNA, se tuvo en cuenta el alto contenido de A+T que presenta este microorganismo, por lo cual las enzimas de elección serían aquellas que reconocieran secuencias con un alto contenido en G+C, existiendo así pocos sitios de corte en el DNA cromosómico de dicha cepa. Las enzimas utilizadas fueron *Eco52I*, *MluI*, *NruI* y *SstI*, siendo los fragmentos utilizados como sondas para los experimentos de hibridación, los mismos que los utilizados para el *Northern-blot*. En la figura 27 se puede observar cómo el patrón de hibridación es el mismo para todos los genes

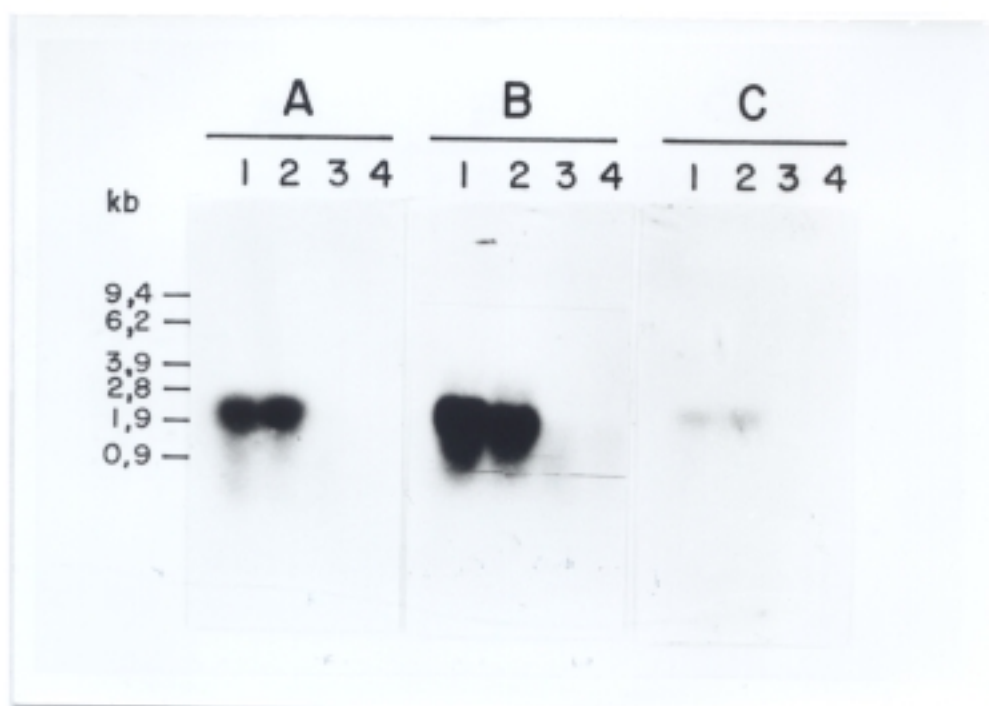


Figura 26- Análisis mediante *Northern-blot* de las unidades de transcripción de los genes *csp*. Calle 1, RNA extraído a una D.O₆₀₀ 0,52 ; Calle 2, RNA extraído a D.O₆₀₀ 1,5; Calle 3, RNA extraído a una D.O₆₀₀ 2,6; Calle 4, RNA extraído a una D.O₆₀₀ 2,7 (fase estacionaria). En el margen izquierdo de la figura se muestran los tamaños en kb de los marcadores de RNA utilizados. A) *Northern-blot* hibridado con el fragmento *PvuII-SspI* del gen *cspA* obtenido del plásmido pKDE543; B) *Northern-blot* hibridado con el fragmento *PstI* del gen *cspC* obtenido del plásmido pKC31. C) *Northern-blot* hibridado con el fragmento *HindIII* del gen *cspB* obtenido del plásmido pKC33. También se hibridó un *Northern-blot* con el fragmento *PstI-EcoRI* del gen *cspD*, perteneciente al plásmido pKC21, pero no se muestra en la figura al no aparecer bandas de hibridación.

a excepción de la digestión *SstI* en donde la banda de hibridación del gen *cspA* es un poco más pequeña. Los tamaños de las bandas de hibridación son de 290 kb para la digestión *Eco52I*, 340 kb para la digestión *MluI*, 350 kb para la digestión *NruI* y 52 kb en el caso del gen *cspA* ó 72 kb para el resto de los genes con la digestión *SstI*. De estos resultados se puede deducir que al menos tres de los genes se localizan en el mismo fragmento de 72 kb del cromosoma, estando juntos los cuatro al menos en un fragmento de 290 kb. Es posible que el fragmento *SstI* de 52 kb se encuentre unido en el cromosoma al fragmento de 72 kb, con lo cual los cuatro genes se localizarían al menos en un fragmento de 124 kb. Por

tanto, parece que los cuatro genes, aunque no constituyan un operón, si parecen localizarse relativamente cerca en el cromosoma bacteriano.

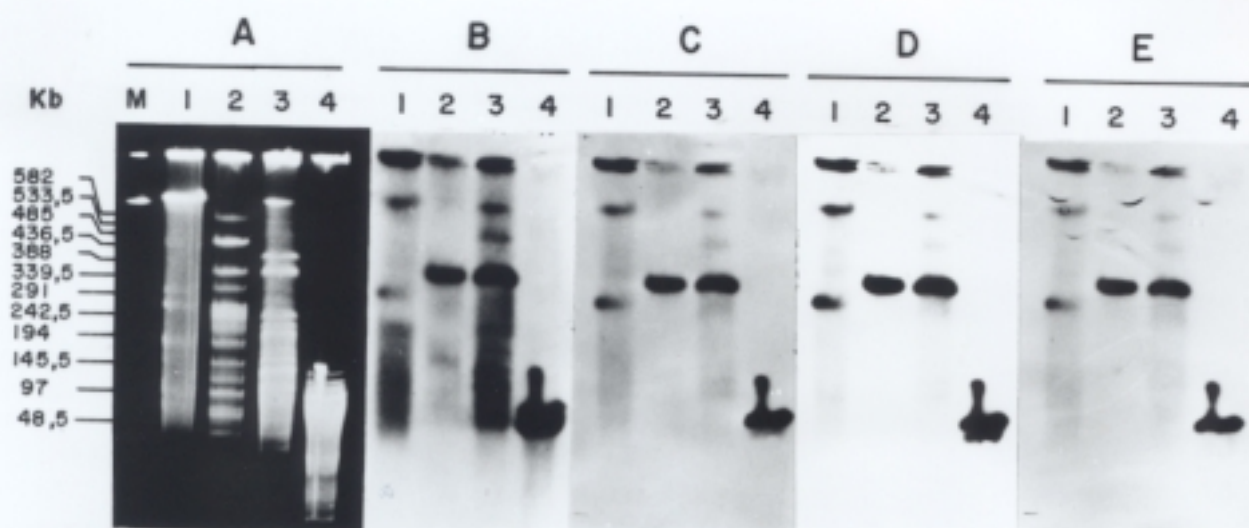


Figura 27- Electroforesis en campo pulsado del DNA de *C. acetobutylicum* NCIB 8052. El panel de la izquierda muestra los fragmentos de restricción separados mediante electroforesis en campo pulsado, teñidos con BrEt y visualizados bajo radiación UV. Las digestiones del DNA se realizaron con las enzimas de restricción *Eco52I* (calle 1), *MluI* (calle 2), *NruI* (calle 3) y *SstI* (calle 4). El panel de la derecha muestra la autorradiografía del *Southern-blot* de dicho gel tras la hibridación con los fragmentos *PvuII-SspI* del plásmido pKDE543 (panel B), *HindIII* del plásmido pKC33 (panel C), *PstI* del plásmido pKC31 (panel D) y *PstI-EcoRI* del plásmido pKC21 (panel E), correspondientes a los genes *cspA*, *cspB*, *cspC* y *cspD* respectivamente. A la izquierda se muestran los marcadores de peso molecular utilizados, expresados en kb (escalera de λ)

Streptococcus pneumoniae

3- CARACTERIZACIÓN MOLECULAR DEL GEN ESTRUCTURAL *pcpA*

No sólo en *S. pneumoniae* existen proteínas con dominios capaces de reconocer e interaccionar con los residuos de colina de los ácidos teicoicos de pared, sino que por los resultados descritos hasta el momento, *C. acetobutylicum* NCIB 8052 posee al menos cuatro genes codificantes para proteínas con dicho dominio. Además, en *C. acetobutylicum* NI-4 y NI-504 existen fragmentos de DNA que hibridan con el fragmento *PvuII-SspI* que codifica para el dominio de unión a colina de la proteína CspA, lo que implica la posible presencia de proteínas de este tipo. En neumococo se habían clonado hasta el momento los genes que codifican para la amidasa LytA (García y cols., 1985) y la proteína de superficie PspA (Yother y Briles, 1992a), aunque se conocía la existencia de otra enzima lítica, la glucosaminidasa (apartado 2.1.2 de la introducción), también dependiente de colina (Sánchez-Puelles y cols., 1986b). El interés por clonar esta última enzima lítica u otras proteínas que contuvieran el dominio de unión a colina, nos llevó a la búsqueda de nuevas estrategias de clonación.

3.1- Hibridación heteróloga del DNA genómico de *S. pneumoniae*

Se realizó un *Southern-blot* con los DNAs cromosómico de la cepa R6 de neumococo y del fago Cp-1, que contiene un gen que codifica para una lisozima dependiente de colina (García y cols., 1987), digiriendo ambos con la enzima de restricción *HincII*. Como DNA control se utilizó el DNA de *C. acetobutylicum* NCIB 8052 digerido con las enzimas de restricción *HincII* y *EcoRV*. La hibridación se realizó en condiciones no estrictas (40°C) utilizando como sonda el fragmento del gen *cspA* que codifica para el dominio de unión a colina (Fig. 28). En las calles correspondientes al DNA de *C. acetobutylicum* aparecen cuatro bandas de hibridación correspondientes a los genes *csp*, detectándose la presencia de dos bandas de hibridación en el DNA de neumococo y una banda en el DNA del fago Cp-1. Esta última corresponde al fragmento *HincII* del DNA del fago Cp-1 que contiene el gen que codifica para la lisozima Cpl1 (López y cols., 1984).

Respecto al perfil de hibridación del DNA de neumococo, conocíamos por secuencia que la banda de hibridación de menor tamaño no podía corresponder al gen *pspA*, aunque sí la mayor (Yother y Briles, 1992a). Para comprobar que ninguna de las dos bandas se

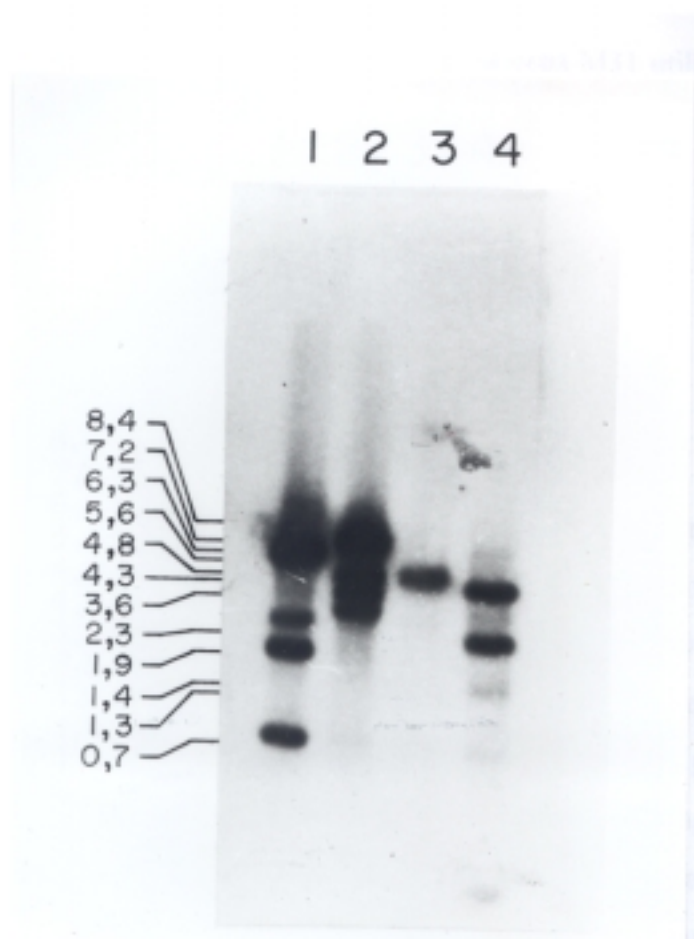


Figura 28- Análisis mediante *Southern-blot* del DNA de *C. acetobutylicum* NCIB 8052, *S. pneumoniae* R6 y fago Cp-1. Calles 1 y 2, DNA de *C. acetobutylicum* NCIB 8052 digerido con *HincII* y *EcoRV*, respectivamente. Calle 3, DNA del fago Cp-1 digerido con *HincII*. Calle 4, DNA de *S. pneumoniae* digerido con *HincII*. En el margen izquierdo de la figura se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (λ BstEII). Se utilizó como sonda el fragmento *PvuII-SspI* del plásmido pKDE543 que codifica para el dominio de unión a colina de la proteína CspA. La temperatura de hibridación y lavado fue de 40 °C.

correspondía con el gen *lytA* se realizaron dos *Southern-blots* incluyendo también el DNA cromosómico de la cepa M31, un mutante de neumococo que carece del gen *lytA* (Sánchez-Puelles y cols., 1986b). Mientras que una membrana se hibridó con la misma sonda y en las mismas condiciones que en el caso anterior, la otra se hibridó en condiciones estrictas (65°) con el plásmido pGL80, el cual contiene el gen *lytA*. En el primer caso, aparecieron dos bandas de hibridación en el DNA de la cepa M31 lo cual quería decir que ninguna de ellas podía corresponder al gen *lytA* (Fig. 29A). Con la segunda hibridación lo

que se trató de demostrar fue la fidelidad del DNA de la cepa M31 utilizado, detectándose la aparición de dos bandas de hibridación con el DNA de la cepa R6 correspondientes a los fragmentos *HincII* que componen el gen *lytA* completo, mientras que no se detectó ninguna banda con el DNA de la cepa M31, como cabía de esperar al carecer del gen *lytA* (Fig. 29B).

3.2- Clonación del gen *pcpA*

Puesto que ninguna de las dos bandas del DNA cromosómico de neumococo que hibridaban con el fragmento que codifica para el dominio de unión a colina de la proteína CspA correspondía al gen *lytA* y, en todo caso, la banda de mayor tamaño podría corresponder al gen *pspA*, cabía la posibilidad de que el otro fragmento que hibridaba contuviera la región que codifica para el dominio de unión a colina de la glucosaminidasa o, al menos, de otra proteína relacionada. Con el fin de clonar los fragmentos de DNA responsables de dichas bandas de hibridación, se decidió construir una genoteca *HincII* con el DNA de la cepa R6 de neumococo utilizando como vector de clonación el plásmido pUC18 digerido con *SmaI* y como cepa receptora *E. coli* TG-1. La selección se realizó utilizando la misma sonda y las mismas condiciones de hibridación del *Southern-blot* descrito en el apartado anterior. De un clon positivo se aisló el plásmido pEK1 con un inserto de 4,4 kb. Al tratarse de un inserto demasiado grande, se decidió subclonar un fragmento más pequeño que contuviera la región responsable de la hibridación pues éste, probablemente, se correspondería con la región codificante del dominio de unión a colina de una nueva proteína. Para ello se realizó un *Southern-blot* de dicho plásmido digerido con diferentes enzimas de restricción e hibridado en las mismas condiciones utilizadas en el análisis de la genoteca (Fig. 30). Puesto que la banda de hibridación de menor tamaño se observó en la digestión *EcoRI*, se procedió a su subclonación en el plásmido pUC18 digerido con *EcoRI* y posterior transformación de *E. coli* TG-1. De esta manera se aisló un nuevo clon positivo que contenía el plásmido recombinante pEK2 con un inserto de 1 kb. Mediante la secuenciación de parte de dicho inserto se pudo comprobar la existencia de secuencias repetidas muy similares a los motivos de unión a colina descritos hasta ahora. Estos datos demostraron que se había conseguido la clonación de un gen codificante para una nueva proteína de neumococo con un dominio de unión a colina.

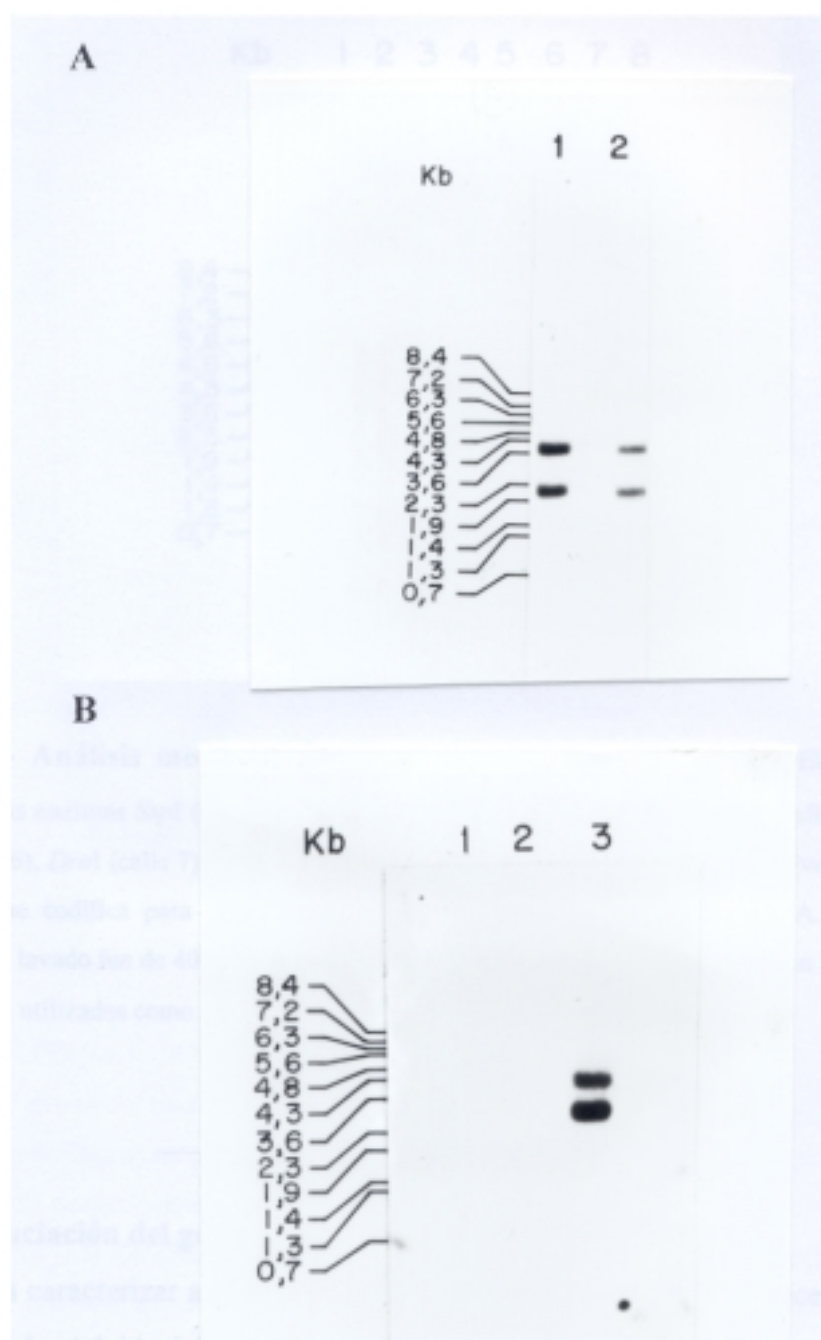


Figura 29-. Análisis mediante *Southern-blot* del DNA de las cepas de neumococo R6 y M31. A) *Southern-blot* hibridado con el fragmento *PvuII-SspI* del plásmido pKDE543 que codifica para el dominio de unión a colina de la proteína CspA. Calle 1, DNA de la cepa de neumococo R6 digerido con la enzima de restricción *HincII*. Calle 2, DNA de la cepa de neumococo M31 digerido con la enzima de restricción *HincII*. La temperatura de hibridación y lavado fue de 40°C. B) *Southern-blot* hibridado con el plásmido pGL80 que contiene el gen *lytA*. Calle 1, DNA de la cepa de neumococo M31. Calle 2, DNA del fago Cp-1. Calle 3, DNA de de la cepa de neumococo R6. La temperatura de hibridación y lavado fue de 65°C. En el margen izquierdo se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (*λ*BsrEII).

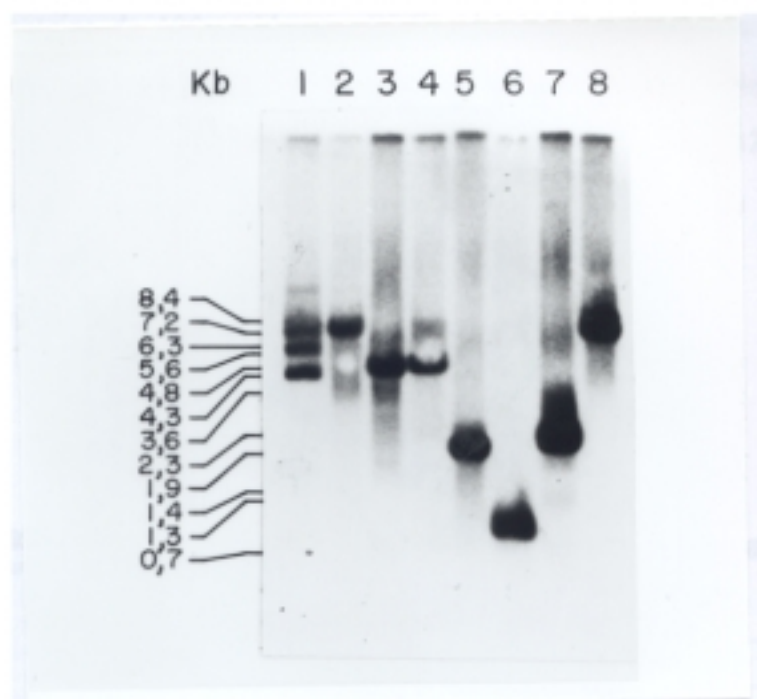


Figura 30- Análisis mediante *Southern-blot* del plásmido pEK1. El plásmido pEK1 se digirió con las enzimas *SspI* (calle 1), *HincII* (calle 2), *HindIII* (calle 3), *EcoRV* (calle 4), *PvuII* (calle 5), *EcoRI* (calle 6), *DraI* (calle 7) y *PstI* (calle 8). Se utilizó como sonda el fragmento *PvuII-SspI* del plásmido pKDE543 que codifica para el dominio de unión a colina de la proteína CspA. La temperatura de hibridación y lavado fue de 40°C. En el margen izquierdo se muestran los tamaños en kb de los fragmentos de restricción utilizados como marcadores (*λBstEII*).

3.3- Secuenciación del gen *pcpA*

Para caracterizar a nivel molecular el nuevo gen clonado, se procedió a la secuencia del inserto de 4,4 kb del plásmido pEK1 utilizando tanto los primers universales como oligonucleótidos sintéticos. En la figura 31 se muestra el mapa de restricción del plásmido pEK1, los subclones preparados para la secuenciación de dicho plásmido y la organización de los posibles genes presentes en el inserto. La secuencia de nucleótidos de las dos cadenas y la secuencia de aminoácidos deducida se muestran en la figura 32. Podemos distinguir cinco marcos de lectura abiertos, ORF1 (del nucleótido 1 al 171) que codifica para una proteína truncada en su extremo 5', ORF2 (del nucleótido 365 al 672) codifica para una posible proteína de 99 aminoácidos, ORF3A (del nucleótido 831 al 1115) codifica para una posible proteína de 99 aminoácidos, ORF3B (del nucleótido 1141 al 1709) codifica para

una posible proteína de 190 residuos y ORF4 (del 2063 al 4190) que codifica para una proteína de 708 residuos. Las ORF1, 2 y 4 se localizan en una de las hebras del DNA situándose la ORF3A y ORF3B en la hebra complementaria (Fig. 31 y 32). Todas las ORFs, a excepción de la ORF1 que está incompleta y la ORF3A, presentan la secuencia de un posible RBS precediendo al codon de iniciación (Murray y Rabinowitz, 1982). Esto sugiere que dichos genes podrían ser traducidos.

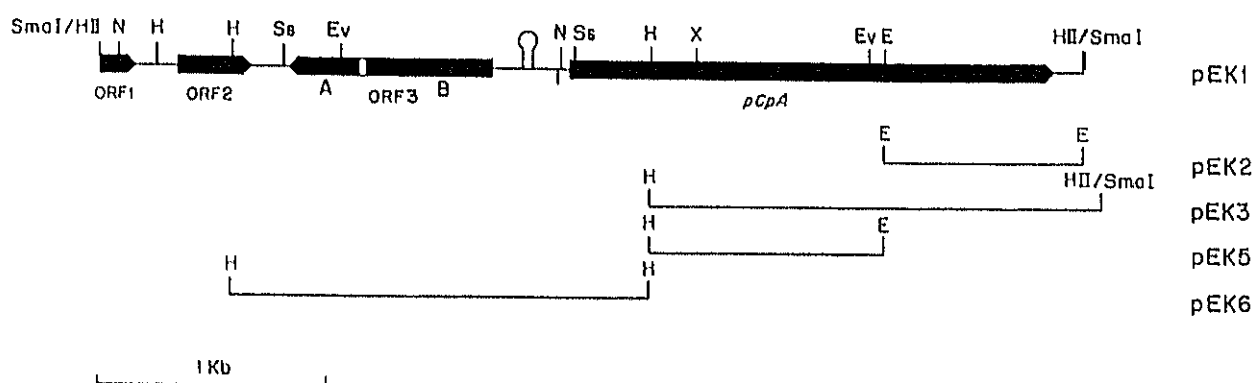


Figura 31- Mapa físico del plásmido pEK1. En la parte superior de la figura se muestra el fragmento de restricción *HincII* del DNA de la cepa de neumococo R6 clonado en el plásmido pEK1. Las flechas negras corresponden a los marcos de lectura abiertos (ORFs), indicándose con la punta de la flecha el sentido de la transcripción. El espacio en blanco presente en la ORF3 hace referencia al salto de fase entre la ORF3A y la ORF3B. Los trazos finos de la parte inferior de la figura representan los subclones derivados del plásmido pEK1 utilizados en la secuenciación. El bucle dibujado entre la ORF3B y el gen *pcpA*, hace referencia a la secuencia palindrónica posiblemente implicada en la regulación del gen *pcpA*.
Dianas de restricción: Ev, *EcoRV*; H, *HindIII*; HII, *HincII*; E, *EcoRI*; N, *NcoI*; Ss, *SspI*.

3.3.1- Análisis de la secuencia de nucleótidos del gen *pcpA*

En la ORF4 se localiza la región que hibrida con el fragmento de DNA que codifica para el dominio de unión a colina del gen *cspA*. Dicho marco de lectura comprende 2.127 nucleótidos que codifican, como ya se dijo anteriormente, para una proteína de 708 residuos que fue denominada PcpA (*Pneumococcal* *C*holine *P*rotein *A*). El codón de iniciación ATG está precedido por la secuencia de un posible RBS (AAGGA, de la posición -8 a la -11).

```

GACGATTTATACGACTTAATACTCTTCGAAAAATCTCTTCAAACCACGTCAGCTTCCATCT
1  -----+-----+-----+-----+-----+ 60
CTGCTAAATATGCTGAATTATGAGAAGCTTTTAGAGAAGTTTGGTGCAGTCGAAGGTAGA
D D L Y D L I L F E N L F K P R Q L P S
ORF1--->

ACAACCTCAAAACAGTGTTTTGAGCAGCCTGCAGCTAGCTTCCTAGTTTGCTCTTTGATT
61  -----+-----+-----+-----+-----+ 120
TGTTGGAGTTTTGTGCACAAAACCTCGTCGGACGTCGATCGAAGGATCAAACGAGAACTAA
T T S K Q C F E Q P A A S F L V C S L I

TTCATTGAGTATAAAAAACAAGAACACCTTGCTTGGCGCAGGGTGTTCGCGTGAAACAGA
121 -----+-----+-----+-----+-----+ 180
AAGTAACTCATATTTTGTCTTGTGGAACGAACCGCTCCACAAAAGCGCACTTTGTCT
F I E Y K N K N T L L G A G C F A *

AGAATTATCTGGTTTCAAATGCTACAGTTAAAATTAGTAAAACAAGAAATAGAAGCTGAA
181 -----+-----+-----+-----+-----+ 240
TCTTAATAGACCAAAGTTTACGATGTCAATTTTAATCATTTTGTCTTTATCTTCGACTT

AAGCCAGCATCTGTAGAAGCTTGGATCATTTCCGTCAAATTAAAAAAGGTTGCTACCGAC
241 -----+-----+-----+-----+-----+ 300
TTCGGTCGTAGACATCTTCGAACCTAGTAAAGGCAGTTTAATTTTTTCCAACGATGGCTG

ATATATAGATTCCAAAAACAAAACGTTAGCGGAACTAGCAGATGTGATTTTATGGAAGTT
301 -----+-----+-----+-----+-----+ 360
TATATATCTAAGGTTTTTGTTTTTGCAATCGCCTTGATCGTCTACACTAAAATACCTCAA

TCGATTTTGCAAATGATCATGCTCATGCATTTTTCATGGATAATGTTGAGTGGAGTCATG
361 -----+-----+-----+-----+-----+ 420
AGCTAAAACGTTTACTAGTACGAGTACGTAAAAAGTACCTATTACAACCTCACCTCAGTAC
1' F A N D H A H A F F M D N V E W S H
ORF2--->

CAGATTCCTTACTTTCTTAGCTTTTGTAGTGACGATGTTGAAGAAGCTTACACAGAAAATG
421 -----+-----+-----+-----+-----+ 480
GTCTAAGAATGAAAGAATCGAAACAATCACTGCTACAACCTTCTTGCAATGTGTCTTTTAC
19' A D S Y F L S F V S D D V E E R Y T E N

TCTATCTGGATAGCCTAAGTGTCAAACAAAAATTTAAGTTTATTTTCGACTTCGGTGATG
481 -----+-----+-----+-----+-----+ 540
AGATAGACCTATCGGATTCACAGTTTGTTTTTTAAATTCAAATAAAGCTGAAGCCACTAC
39' V Y L D S L S V K Q K F K F I F D F G D

AATGGCGTTTTGAATGCCAAGTGCTGAGAGAAATCGAGACAGAGGACGAAGAAGCTTATC
541 -----+-----+-----+-----+-----+ 600
TTACCGCAAACTTACGGTTCACGACTCTCTTTAGCTCTGTCTCCTGCTTCTTCGAATAG
59' E W R F E C Q V L R E I E T E D E E A Y

TCGTACGTTTCGGTTGGAACGTCGCCAGAACAATATCCAGATTATGATGGTTTTGACTATG
601 -----+-----+-----+-----+-----+ 660
AGCATGCAAGCCAACCTTGCAGCGGTCTTGTATAGGTCTAATACTACCAAACTGATAC
79' L V R S V G T S P E Q Y P D Y D G F D Y

AAGAATGGTAAAAATTGAAATCAGTCTGTGTAGGCTTAGTATTTCAATAGACTTCCTGCAA
661 -----+-----+-----+-----+-----+ 720
TTCTTACCATTTTAACTTTAGTCAGACACATCCGAATCATAAAGTTATCTGAAGGACGTT
99' E E W *

AACTAGAATCCTAGTTTCATGATTGATAATACCAGCACTCAAATTCATTTCGTAATCCGAAG
721 -----+-----+-----+-----+-----+ 780
TTGATCTTAGGATCAAGTACTAACTATTATGGTCGTGAGTTTAAGTAAGCATTAGGCTTC

CGTTTACGATGATTTTCGATAGGTTGTTGAAAATATTTTAAACGTTTTTAGTTTGGCAAAG
781 -----+-----+-----+-----+-----+ 840
GCAAATGCTACTAAAGCTATCCAACAACCTTTTATAAAATTTGCAAAAATCAAACCGTTTC
|----->| -- -- --> * N P L S 91'
RD RI
ATGTTCTCAACCTTGCTTCTTCTTGATAGACGCATGGTTATAGGCTTTATTTTCAGCTG
841 -----+-----+-----+-----+-----+ 900
TACAAGAGTTGGAACGAAGAAGAACTATCTGCGTACCAATATCCGAAATAAAGTCGAC
T R L R A E K K I S A H N Y A K N E A T 71'

```

```

TTAGCGGCTTGAGTTTGCTGGATTACGTGGAGTTTGTACTTGAGGATATATCTTCATGA
901  -----+-----+-----+-----+-----+ 960
AATCGCCGAACCTCAAACGACCTAAATGCACCTCAAACATGAACCTCTATATAGAAGTACT
  L P K L K S S K R P T Q V Q P Y I K M L 51'

GCCCTTGATAACCACTGTCAGCCAAGATTTTACCAGCTTGTTTCGATATTTCTACGACTCA
961  -----+-----+-----+-----+-----+ 1020
CGGGAACCTATTGGTGACAGTCGGTTCTAAATGGTCGAACAAGCTATAAAGATGCTGAGT
  G Q Y G S D A L I K G A Q E I N R R S M 31'

TTTTGAACAACCTTCATATCATGACTATAGTTTACAGCGATATCCAAAGAAACAATTCTCC
1021 -----+-----+-----+-----+-----+ 1080
AAAACCTTGTGAGTATAGTACTGATATCAAGTGTCTGCTATAGGTTTCTTTGTTAAGAGG
  K F L K M D H S Y N V A I D L S V I R G 11'

CTTGACTTGTGACAATCGCTTGAGCCTTCATAGCGTGAAATTTCTTTTTTACCAGAATTAT
1081 -----+-----+-----+-----+-----+ 1140
GAACTGAACACTGTTAGCGAAGCTCGAAGTATCGCACTTTAAAGAAAAATGGTCTTAATA
  Q S T V I A Q A K M A H F K K K G S N +1 1'
<---ORF3A

TCGCTAATTCCTTTTTTTAGGGCGATTGATTTTTTACTTCCGTCGCATCAATAATCATTACC
1141 -----+-----+-----+-----+-----+ 1200
AGCGATTAAAGAAAAAATCCCGCTAACTAAAAATGAAGGCAGCGTAGTTATTAGTAATGG
  E S I R K K P R N I K V E T A D I I M V 170'

GTGTCCTCAGAACTGAGAGGAGTTCTTGAAATCGTAACACCACTTTGAACAAGAGTTACT
1201 -----+-----+-----+-----+-----+ 1260
CACAGGAGTCTTGACTCTCCTCAAGAACTTTAGCATTGTGGTGAACTTTGTTCTCAATGA
  T D E S S L P T R S I T V G S Q V L T V 150'

TCAACCCATTGGCTCCGACGGATTAAAGTTGCTTTCGTGAATACCAAAAATCAGCCGCAAT
1261 -----+-----+-----+-----+-----+ 1320
AGTTGGGTAACCGAGGCTGCCTAATTCACGAAAGCACTTATGGTTTTTGTAGTCGGCGTTA
  E V W Q S R R I L N S E H I G F I L R L 130'

TTGTTTCATAAGTTTCGATATTCTCGCACATATTGAAGAGTGGCCATAAGAAGGCTTCTAGG
1321 -----+-----+-----+-----+-----+ 1380
AACAAGTATTCAAGCTATAAGAGCGGTGATAACTTCTCACCAGTATTCCTCCGAAGATCC
  K N M L E I N E C M N F L P W L F A E L 110'

CTTAATTTAGGTTTTTCGTCCACCTTTTGCCTGTTTAAAGTTGATAAGCTGTTTTTAATACA
1381 -----+-----+-----+-----+-----+ 1440
GAATTTAAATCCAAAAGCAGGTGGAAAACGCACAAATTCAACTATTCGACAAAAATTATGT
  S L K P K R G G K A H K L Q Y A T K L V 90'

GCTAACATCTCTTTCAAAGTCGTGCGTTGAACACCAACAAGACGCTTAAATCGTGCATCG
1441 -----+-----+-----+-----+-----+ 1500
CGATTGTAGAGAAGTTTTTACGACGCAACTTGTGGTTGTTCTGCGAATTTAGCACGTAGC
  A L M E E F T T R Q V G V L R K F R A D 70'

GTTAGTTGTTTAGTTGCTTCATAATTTTGTTCGCGAGGAGTCTATTGACTCTTTGGTAG
1501 -----+-----+-----+-----+-----+ 1560
CAATCAACAAATCAACGAAGTATTAACAAAGCGTCCCTCAGATAACTGAGAAACCATC
  T L Q K T A E Y N Q K A P L R N V R Q Y 50'

GTGTCAATCTTTTTTTCATCTATCCCGAGAATTATTTCCCGCCATTTGTATTTGCAAAT
1561 -----+-----+-----+-----+-----+ 1620
CACAGTTAGAAAAAAGTAGATAGGGCTCTTAATAAAGGGCGGTAAACATAACGTTTA
  T D I K K E D I G L I I K G G N T N A F 30'

GCTGAGTAGGTTTCCAGAAAGACTCTGGAAGATTGTTTTTAGCTTTTTTGTATTCTAAA
1621 -----+-----+-----+-----+-----+ 1680
CGACTCATCCAAAGGGTCTTTCTGAGACCTTCTAACAAAAATCGAAAAACATAAGATTT
  A S Y T E W F S E P L N N K A K K Y E L 10'

TCAACCCCTTCAAATTTTAAAGTCCATATTTTCCCTTACATCTGAAAAAGTGGTTCTGG
1681 -----+-----+-----+-----+-----+ 1740
AGTTGGGGAAGTTTAAATTCAGGTATAAAAGGAAATGTAGACTTTTTTACCAAGACC
  D V G E F K L D M -10 1'
<---ORF3B

```

```

TATTTGTTCAAGTTGAGTGATAATATAGCGAATTGAATTCGAGAGTTTTTACTCAGTTA
1741 -----+-----+-----+-----+-----+ 1800
ATAAACAAAGTTCAACTCACTATTATATCGCTTAACTTAAAGCTCTCAAAAATGAGTCAAT
      -35          <-- - - --      ||----->
                      RI          RD          -

ATTTCTTTTAAACCCACTTTAATTTCTTTTAAACCCACTTTAATTTCTTTTAAACAC
1801 -----+-----+-----+-----+-----+ 1860
TAAAGAAAAAATTGGGTGAAATTAAAGAAAAAATTGGGTGAAATTAAAGAAAAAATTGTG
----->----->----->----->----->

                                -35          -10
GGGTTAAAAAAGAAATTAAAGTGGGTAAATTTTCCTTGACTTAAATTTAAAAATGATAT
1861 -----+-----+-----+-----+-----+ 1920
CCCAATTTTTTCTTTAATTTACCCAATTAAAGGAAGTGAATTTTAAATTTTACTATA
      +1
ACTGTGTTTATGGAGAGATGACAATTATAAAATGTACTACTCTATTCTACTAGATAATG
1921 -----+-----+-----+-----+-----+ 1980
TGACACAAATACCTCTCTACTGTTAAATATTTTACATGATGAGATAAGATGATCTATTAC

ATTGAATAAACTTTTTATAACAACGGCTAGCAAAGTTATAATTTTATATCTATTAGAAA
1981 -----+-----+-----+-----+-----+ 2040
TAACTTATTTTGAAAAATATTGTTGCCGATCGTTTCAATATTAATAATATAGATAATCTTT

ATAAATATGTAAGGAAATTTAAATGAAAAAACTACAATATTATCATTAACACAGCTGC
2041 -----+-----+-----+-----+-----+ 2100
TATTTATACATTCCTTTAAATTTACTTTTTTTGATGTTATAATAGTAATTGATGTCGACG
1'          M K K T T I L S L T T A A
          ORF4--->

GGTTATTTTAGCAGCATATGTCCCTAATGAACCAATCCTAGCAGCATATGTCCCTAATGA
2101 -----+-----+-----+-----+-----+ 2160
CCAATAAAATCGTCGTATACAGGGATTACTTGGTTAGGATCGTCGTATACAGGGATTACT
14'  V I L A A Y V P N E P I L A A Y V P N E

ACCAATCCTAGCAGATACTCCTAGTTCGGAAGTAATCAAAGAGACTAAAGTTGGAAGTAT
2161 -----+-----+-----+-----+-----+ 2220
TGGTTAGGATCGTCTATGAGGATCAAGCCTTCATTAGTTTCTCTGATTCAACCTTCATA
34'  P I L A D T P S S E V I K E T K V G S I

TATTC AACAAAATAATATCAAATATAAGGTTCTAACTGTAGAAGGTAACATAGGAACTGT
2221 -----+-----+-----+-----+-----+ 2280
ATAAGTTGTTTATTATAGTTTATATTCCAAGATTGACATCTTCCATTGTATCCTTGACA
54'  I Q Q N N I K Y K V L T V E G N I G T V

TCAAGTGGGTAAATGGAGTTACTCCTGTAGAGTTTGAAGCTGGTCAAGATGGAAGACATT
2281 -----+-----+-----+-----+-----+ 2340
AGTTCACCCATTACCTCAATGAGGACATCTCAAACCTCGACCAGTTCTACCTTTTGGTAA
74'  Q V G N G V T P V E F E A G Q D G K P F

CACGATTCCTACAAAAATCACAGTAGGTGATAAAGTATTTACCGTTACTGAAGTAGCTAG
2341 -----+-----+-----+-----+-----+ 2400
GTGCTAAGGATGTTTTTATGTGTCATCCACTATTTTATAAATGGCAATGACTTCATCGATC
94'  T I P T K I T V G D K V F T V T E V A S

TCAAGCTTTTAGTTATTATCCAGATGAAACAGGTAGAATTGTCTACTATCCTAGCTCTAT
2401 -----+-----+-----+-----+-----+ 2460
AGTTCGAAAATCAATAATAGGTCTACTTTGTCCATCTTAACAGATGATAGGATCGAGATA
114' Q A F S Y Y P D E T G R I V Y Y P S S I

TACTATCCCATCAAGCATAAAAAAATACAAAAAAGGCTTCATGGAAGTAAAGCTAA
2461 -----+-----+-----+-----+-----+ 2520
ATGATAGGGTAGTTTCGTATTTTTTTTATGTTTTTTTCCGAAGGTACCTTCATTTCGATT
134' T I P S S I K K I Q K K G F H G S K A K

AACTATTATTTTTGACAAAGGCAGTCAGCTGGAGAAAATTGAAGATAGAGCTTTTGATTT
2521 -----+-----+-----+-----+-----+ 2580
TTGATAATAAAACTGTTTCCGTCAGTCGACCTCTTTAACTTCTATCTCGAAAACATAA
154' T I I F D K G S Q L E K I E D R A F D F

```

TTCTGAATTAGAAGAGATTGAATTGCCTGCATCTCTAGAATATATTGGAACAAGTGCATT
 2581 -----+-----+-----+-----+-----+ 2640
 AAGACTTAATCTTCTCTAACTTAACGGACGTAGAGATCTTATATAACCTTGTTACGTAA
 174' S E L E E I E L P A S L E Y I G T S A F

 TTCTTTTAGTCAAAAATTGAAAAAGCTAACCTTTTCCTCAAGTTCAAAATTAGAATTAAT
 2641 -----+-----+-----+-----+-----+ 2700
 AAGAAAATCAGTTTTTAACTTTTTTCGATTGGAAAAGGAGTTCAAGTTTTAATCTTAATTA
 194' S F S Q K L K K L T F S S S S K L E L I

 ATCACATGAGGCTTTTGCTAATTTATCAAATTTAGAGAACTAACATTACCAAAATCGGT
 2701 -----+-----+-----+-----+-----+ 2760
 TAGTGTACTCCGAAAACGATTAAATAGTTTAAATCTCTTTGATTGTAATGGTTTTAGCCA
 214' S H E A F A N L S N L E K L T L P K S V

 TAAACATTAGGAAGTAATCTATTTAGACTCACTACTAGCTTAAACATGTTGATGTTGAG
 2761 -----+-----+-----+-----+-----+ 2820
 ATTTTGTAAATCCTTCATTAGATAAATCTGAGTGATGATCGAATTTGTACAACACTACAAC
 234' K T L G S N L F R L T T S L N M L M L R

 AGGAATGATCGTTGCCTCAGTTGATGGTGTTCGTTTCAAAGTAAAACTCAATTAATTTA
 2821 -----+-----+-----+-----+-----+ 2880
 TCCTTACTAGCAACGGAGTCAACTACCACAAAGCAAAGTTTCATTTTGTAGTTAATTAAT
 254' G M I V A S V D G V S F Q S K T Q L I Y

 TTATCCAAGTCAAAAAAATGACGAAAGTTATAAAACGCCTAAGGAGACAAAAGAACTTGC
 2881 -----+-----+-----+-----+-----+ 2940
 AATAGGTTTCAGTTTTTTTACTGCTTTCAATATTTTGCAGATTCTCTGTTTCTTGAACG
 274' Y P S Q K N D E S Y K T P K E T K E L A

 ATCATATTCGTTTAAATAAAAATTCCTTACTTGAAAAAAGTTCGAATTGAATGAAGGTTTACA
 2941 -----+-----+-----+-----+-----+ 3000
 TAGTATAAGCAATATTTTTTAAAGAAAGAACTTTTTTGAAGCTTAACCTTACCTCAAATGT
 294' S Y S F N K N S Y L K K L E L N E G L Q

 AAAAATCGGTACTTTTGCATTTGCGGATGCGACCAAAGTGAAGAAATTAGCTTACCAAA
 3001 -----+-----+-----+-----+-----+ 3060
 TTTTTAGCCATGAAAACGTAACGCCTACGCTGGTTTGAAGTCTTTTAAATCGAATGGTTT
 314' K I G T F A F A D A T K L E E I S L P N

 TAGTTTAGAAACTATTGAACGTTTAGCCTTTTACGGTAATTTAGAATTAAAAGAACTTAT
 3061 -----+-----+-----+-----+-----+ 3120
 ATCAAATCTTTGATAACTTGCAAAATCGGAAATGCCATTAAATCTTAATTTCTTGAATA
 334' S L E T I E R L A F Y G N L E L K E L I

 ATTACCAGATAATGTTAAAAATTTTGGTAAACACGTTATGAACGGTTTACCAAAATTTTT
 3121 -----+-----+-----+-----+-----+ 3180
 TAATGGTCTATTACAATTTTTAAACCATTTGTGCAATACTTGCCAAATGGTTTTTAAAAA
 354' L P D N V K N F G K H V M N G L P K F L

 AACATTATCTGGTAATAATATCAACTCATTGCCGTCCTTCTTCCTAAGTGGCGTCTTAGA
 3181 -----+-----+-----+-----+-----+ 3240
 TTGTAATAGACCATTATTATAGTTGAGTAACGGCAGGAAGGATTCACCGCAGAATCT
 374' T L S G N N I N S L P S F F L S G V L D

 TTCATTAAAGGAAATTCATATTAAGAATAAAAGTACAGAGTTTTCTGTGAAAAAAGATAC
 3241 -----+-----+-----+-----+-----+ 3300
 AAGTAATTTCCCTTAAAGTATAATCTTATTTTCATGTCTCAAAGACACTTTTTTCTATG
 394' S L K E I H I K N K S T E F S V K K D T

 ATTTGCAATTCCTGAAACTGTTAAGTTCTATGTAACATCAGAACATATAAAAGATGTTCT
 3301 -----+-----+-----+-----+-----+ 3360
 TAAACGTTAAGGACTTTGACAAATCAAGATACATTGTAGTCTTGTAATTTTCTACAAGA
 414' F A I P E T V K F Y V T S E H I K D V L

 TAAATCAAATTTATCTACTAGTAATGATATCATTTGTTGAAAAAGTAGATAATATAAAACA
 3361 -----+-----+-----+-----+-----+ 3420
 ATTTAGTTTAAATAGATGATCATTACTATAGTAACAACTTTTTCATCTATTATTTTGT
 434' K S N L S T S N D I I V E K V D N I K Q

```

AGAAACTGATGTAGCTAAACCTAAAAAGAATTCTAATCAGGGAGTAGTTGGTTGGGTAA
3421 -----+-----+-----+-----+-----+-----+-----+ 3480
TCTTTGACTACATCGATTGGATTTTCTTAAAGATTAGTCCCTCATCAACCAACCCAATT
454' E T D V A K P K K N S N Q G V V G W V K

AGACAAAGGTTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3481 -----+-----+-----+-----+-----+-----+-----+ 3540
TCTGTTTCCAAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
474' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGTTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3541 -----+-----+-----+-----+-----+-----+-----+ 3600
TCTGTTTCCAAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
494' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGTTTATGGTATTACTTAAATGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3601 -----+-----+-----+-----+-----+-----+-----+ 3660
TCTGTTTCCAAATACCATAATGAATTTACTTAGTCCAAGTTACCGATGACCAACCCAATT
514' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3661 -----+-----+-----+-----+-----+-----+-----+ 3720
TCTGTTTCCGAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
534' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3721 -----+-----+-----+-----+-----+-----+-----+ 3780
TCTGTTTCCGAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
554' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAATGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3781 -----+-----+-----+-----+-----+-----+-----+ 3840
TCTGTTTCCGAATACCATAATGAATTTACTTAGTCCAAGTTACCGATGACCAACCCAATT
574' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3841 -----+-----+-----+-----+-----+-----+-----+ 3900
TCTGTTTCCGAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
594' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAATGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3901 -----+-----+-----+-----+-----+-----+-----+ 3960
TCTGTTTCCGAATACCATAATGAATTTACTTAGTCCAAGTTACCGATGACCAACCCAATT
614' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
3961 -----+-----+-----+-----+-----+-----+-----+ 4020
TCTGTTTCCGAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
634' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAACGAATCAGGTTCAATGGCTACTGGTTGGGTAA
4021 -----+-----+-----+-----+-----+-----+-----+ 4080
TCTGTTTCCGAATACCATAATGAATTTGCTTAGTCCAAGTTACCGATGACCAACCCAATT
654' D K G L W Y Y L N E S G S M A T G W V K

AGACAAAGGCTTATGGTATTACTTAAATGAATCAGGTTCAATGGCTACTGGTTGGGTAA
4081 -----+-----+-----+-----+-----+-----+-----+ 4140
TCTGTTTCCGAATACCATAATGAATTTACTTAGTCCAAGTTACCGATGACCAACCAAATT
674' D K G L W Y Y L N E S G S M A T G W F K

AGTTTCTGGTAAATGGTACTATACCTATAATTCAGGAGATTTATTTAGTAAACTCGACT
4141 -----+-----+-----+-----+-----+-----+-----+ 4200
TCAAAGACCATTATCCATGATATGGATATTAAGTCTCTAAAATAAATCATTTGAGCTGA
694' V S G K W Y Y T Y N S G D F I *

ACACCCGATGGCTATCGAGTCAATGCTACGGTGAGTGGGTAGATAGGTATCCATTGCTAA
4201 -----+-----+-----+-----+-----+-----+-----+ 4260
TGTGGGCTACCGATAGCTCAGTTACGATGCCACTCACCCATCTATCCATAGGTAACGATT

```

```

      ATCTAATCATTCTGAACTAAAAAATCTTCTTACATATGTAGGAAGATTTTTGGATTCTC
4261 -----+-----+-----+-----+-----+-----+ 4320
      TAGATTAGTAAGACTTGATTTTTTAGAAGAATGTATACATCCTTCTAAAAAACCTAAGAG

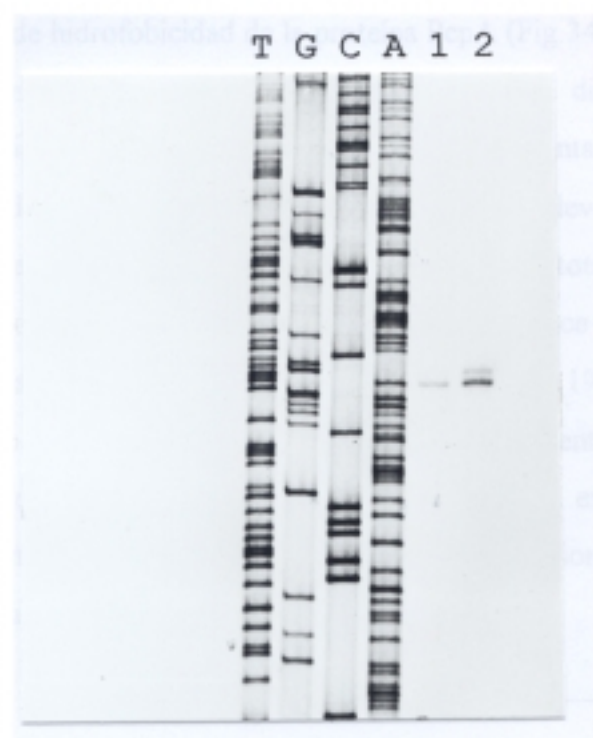
      AAAACTAGTTTAAAAGACTACGCTATTTAAACCAATTTTAAGAATGGATATATAAGGATA
4321 -----+-----+-----+-----+-----+-----+ 4380
      TTTTGATCAAATTTTCTGATGCGATAAATTGGTTAAAATTCTTACCTATATATTCCTAT

      TTCAGAAATCTAGTTGGTC
4381 -----+-----4399
      AAGTCTTTAGATCAACCAG

```

Figura 32- Secuenciación del inserto de 4,4 kb del plásmido pEK1. Se muestran la secuencia de nucleótidos de las dos cadenas así como la secuencia de aminoácidos de las posibles ORFs. Con un asterisco se señalan los codones de terminación de la traducción. Los nucleótidos marcados en negrita corresponden a los posibles RBSs y regiones promotoras -10 y -35. Los nucleótidos subrayados hacen referencia a la presencia de posibles bucles de terminación. En la región anterior a la ORF4, en negrita y subrayado, se muestra la secuencia palindrómica posiblemente implicada en la regulación de la transcripción del gen *pcpA*. Solapando con esta secuencia, se representan con flechas (--->), la presencia de repeticiones directas. El nucleótido marcado en negrita (+1), representa el sitio de iniciación de la transcripción del gen *pcpA*. Las flechas (|--->) representan las posibles repeticiones directas (RD) e invertidas (RI), presentes en la secuencia de inserción IS1381. Los aminoácidos marcados en negrita, son resultantes del solapamiento de la ORF3A y ORF3B, teniendo en cuenta un salto de fase en el marco de lectura de la ORF3B, de tipo +1. .

Se ha determinado mediante la técnica de *primer extensión*, el sitio de iniciación de la transcripción. La transcripción del gen *pcpA* parece comenzar en la adenina situada a 132 pb del codon de iniciación (Fig. 33). Precediendo al sitio de iniciación de la transcripción se localizan dos secuencias características de promotores de genes procariotas reconocidas por la subunidad σ -70 de la RNA polimerasa de *E. coli* (Fig. 33). Es curioso observar cómo existe una gran distancia entre el sitio de iniciación de la transcripción y el ATG, al igual que ocurría con el gen *lytA*. Aunque ya se han descrito en la literatura regiones *leader* similares e incluso de mayor longitud, se trata de una propiedad poco frecuente y generalmente relacionada con mecanismos de regulación génica (Friesen y Tropak, 1983; Benner-Lugger y Boss, 1988). Otro aspecto interesante es la existencia de una estructura en bucle que se localiza 6 pb por delante de la caja -35, con una energía libre teórica de -41,2 kcal/mol y que podría estar implicada en la regulación transcripcional del gen *pcpA*. A 87 nucleótidos del codón de terminación TAG, se localiza una secuencia palindrómica capaz de formar una estructura en bucle con una energía libre teórica de -14,9 kcal/mol, que podría actuar como un terminador de tipo rho-independiente.



```

-----
ACTTTAATTTCTTTTTTAACCCACTTTAATTTCTTTTTTAACACGGGTAAAAAAGAAATTAAGTGGGTTAATT
      -35              -10      +1---->
TTCCTTGACTTAAAAATTTAAAAATGATATACTGTGTTTATGGAGAGATGACAATTTATAAAATGTACTACTCTAT
TCTACTAGATAATGATTGAATAAACTTTTTATAACAACGGCTAGCAAAGTTATAATTTTATATCTATTAGAAAA
TAAATATGTAAGGAAATTTAAATGAAAAAACTACAATATTATCATTAACTACAGCTGCGGTTATTTTAGCAGCA
      M K K T T I L S L T T A A V I L A A

```

Figura 33- Análisis mediante *primer extensión* del sitio de iniciación de la transcripción del gen *pcpA*. En la parte superior izquierda se muestra parte de la secuencia del plásmido pEK1 realizada con el oligo OEK5 (5'-ATAACCGCAGCTGTAG-3'). En la parte superior derecha se muestran los sitios de iniciación de la transcripción de dos muestras de RNA extraídas a una densidad óptica de 344 N (calle 1) y 800 N (calle 2). En la parte inferior de la figura se muestra un fragmento de la secuencia situada delante del gen *pcpA*, marcándose con +1 el sitio de iniciación de la transcripción. Más arriba se localiza las cajas -10 y -35 de la región promotora y por detrás en negrita el RBS y el codón de iniciación de la proteína PcpA. Las flechas hacen referencia a la secuencia palindrómica posiblemente implicada en regulación.

3.3.2- Análisis de la secuencia de aminoácidos

La proteína PcpA presenta un peso molecular de 79 kDa, con un pI de 9,3. Al observar el perfil de hidrofobicidad de la proteína PcpA (Fig. 34) se puede comprobar cómo la proteína está estructurada en dos dominios claramente diferenciados. El dominio N-terminal, constituido por los 469 primeros residuos, presenta un perfil de hidrofobicidad heterogéneo con dos picos hidrofóbicos y dos hidrofílicos relevantes. El dominio C-terminal constituido por los restantes aminoácidos presenta un perfil totalmente homogéneo. Esto es debido a que el dominio C-terminal está estructurado en doce motivos repetidos idénticos, de 20 aminoácidos de longitud, que terminan en una cola de 19 aminoácidos que representa posiblemente otro motivo degenerado (Fig. 35A). Dicha identidad no se manifiesta sólo a nivel de aminoácidos sino también a nivel de nucleótidos, a excepción de los aminoácidos asparagina y glicina que en 4 y 5 ocasiones respectivamente son codificados por los tripletes GGT y AAT en lugar de por GGC y AAC (Fig. 35A).

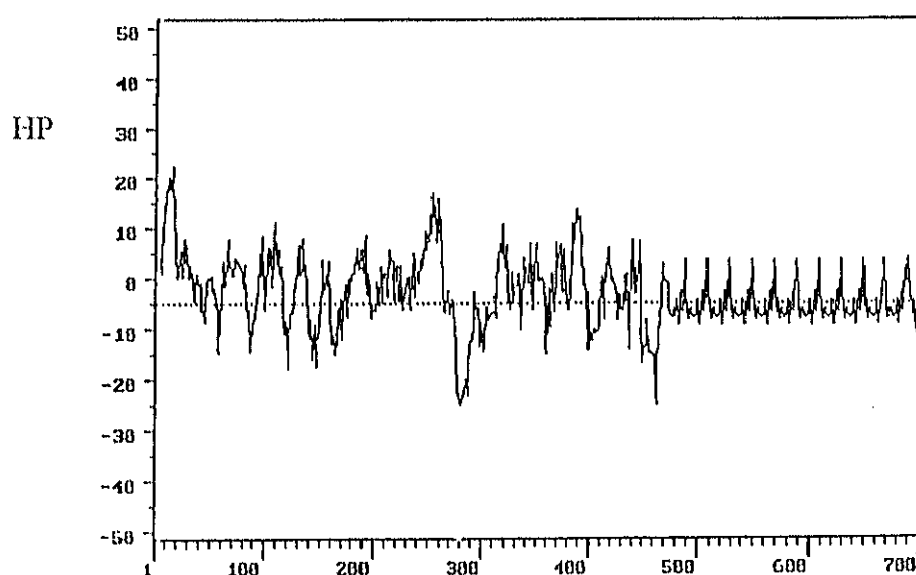


Figura 34- Perfil de hidrofobicidad de la proteína PcpA. Se llevó a cabo de acuerdo con el método de Kyte y Doolittle (1982).

A

G	W	V	K	D	K	G	L	W	Y	Y	L	N	E	S	G	S	M	A	T
GGT	TGG	GTT	AAA	GAC	AAA	GGT	TTA	TGG	TAT	TAC	TTA	AAC	GAA	TCA	GGT	TCA	ATG	GCT	ACT
						C							T						

B

I	15-24	I L A A Y V P N E P
II	25-34	I L A A Y V P N E P
III	35-37	I L A

C

			5	10	15	20	25																					
I	143-166	Q	K	K	G	F	H	G	S	-	K	-	A	K	T	I	I	F	D	K	G	S	Q	L	E	K	I	
II	167-188	E	D	R	A	F	D	F	S	E	-	-	L	E	E	I	E	L	P	A	S	-	-	L	E	Y	I	
III	189-213	G	T	S	A	F	S	F	S	Q	K	-	L	K	K	L	T	F	S	S	S	S	S	K	L	E	L	I
IV	214-236	S	H	E	A	F	A	N	L	S	N	-	L	E	K	L	T	L	P	K	S	-	-	V	K	T	L	
V	237-260	G	S	N	L	F	R	L	T	T	S	-	L	N	M	L	M	L	R	G	M	I	-	V	A	S	V	
VI	293-315	A	S	Y	S	F	N	K	N	S	Y	-	L	K	K	L	E	L	-	-	N	E	G	L	Q	K	I	
VII	316-338	G	T	F	A	F	A	D	A	T	K	-	L	E	E	I	S	L	P	-	N	S	-	L	E	T	I	
VIII	339-361	E	R	L	A	F	Y	G	N	L	E	-	L	K	E	L	I	L	P	D	N	-	-	V	K	N	F	
IX	362-383	G	K	H	V	M	N	G	L	P	K	-	F	-	-	L	T	L	S	G	N	N	-	I	N	S	L	
X	384-409	P	S	F	F	L	S	G	V	L	D	S	L	K	E	I	H	I	K	N	K	S	T	E	F	S	V	
Consenso		x	x	x	A	F	x	x	x	x	x	-	L	x	x	L	x	L	x	x	x	-	x	L	x	x	I	

Figura 35- Secuencias de los motivos repetidos presentes en la proteína PcpA A)

Secuencia de nucleótidos y aminoácidos de los motivos de unión a colina presentes en el dominio C-terminal de la proteína PcpA. Aunque los 12 motivos son idénticos a nivel de aminoácidos, en negrita se marcan las únicas variaciones a nivel de nucleótidos. B) Secuencia de los motivos repetidos presentes en el extremo N-terminal, los cuales son idénticos a nivel de aminoácidos pero no de nucleótidos. C) Secuencias de los 10 motivos ricos en leucina y otros aminoácidos no polares, presentes en el dominio N-terminal de la proteína PcpA. En negrita se marcan los aminoácidos conservados en, al menos, un 50% de los motivos.

Si se analiza detenidamente la estructura primaria de la región N-terminal de esta proteína se pueden distinguir dos tipos de repeticiones. En el extremo de esta región se localizan dos motivos completos y otro truncado, de 10 aminoácidos de longitud, que son idénticos en la composición de aminoácidos pero no en la secuencia de nucleótidos (Fig. 35B). A partir del residuo 143 se localiza una región con 10 motivos de 22 a 26 aminoácidos de longitud, ricos en aminoácidos no polares, predominantemente leucina aunque también son abundantes la isoleucina, valina y fenilalanina, alternando con aminoácidos marcadamente hidrofílicos, lo que confiere a estas repeticiones un carácter anfipático (Fig. 35C). Estos motivos se agrupan en dos bloques de cinco repeticiones cada uno, comprendiendo el primer bloque desde el residuo 143 al 260 y el segundo bloque

desde el residuo 293 al 409 (Fig. 35C) unidos por un péptido de interconexión con carácter hidrofílico. Estos motivos repetidos constituyen más del 50% del extremo N-terminal de la proteína y aunque el resto de su secuencia presenta un alto contenido en aminoácidos no polares, no parece evidente que estén agrupados constituyendo motivos como los descritos anteriormente.

Considerando la secuencia consenso de estas repeticiones hay que destacar los residuos de leucinas conservados en las posiciones 12, 15 y 17 y que están presentes, al menos, en el 60% de los motivos repetidos. En algunos motivos estas leucinas consenso son sustituidas por los aminoácidos isoleucina, fenilalanina o alanina. Hay que destacar que tanto la leucina en la posición 12, como la fenilalanina en la posición 5, son los aminoácidos más conservados, estando presentes en el 80% de los motivos. Otras posiciones conservadas con aminoácidos hidrofóbicos son la 23 y la 26. La primera presenta un 50% de leucinas sustituidas en el resto de las repeticiones mayoritariamente por valina. La segunda, sin embargo, presenta un 50% de isoleucinas sustituidas en los otros motivos por aminoácidos hidrofóbicos como leucina, valina ó fenilalanina. La posición 4 con un 50% de alanina, y la 20 que está ocupada en la mayor parte de los casos por serina o asparagina, son también dignas de mención.

3.4- Comparación de la proteína PcpA con otras proteínas del banco de datos

Al comparar la secuencia de aminoácidos del dominio C-terminal, con las secuencias presentes en la base de datos del EMBL, se comprobó que la región C-terminal presentaba una gran similitud con los dominios de unión a colina ya descritos previamente (Fig. 57, pag. 162).

El dominio N-terminal mostró una elevada similitud con los receptores de las hormonas gonadotrópicas así como con la chaoptina, una glicoproteína de superficie presente en las células fotorreceptoras del ojo compuesto de la mosca. En la figura 36A se compara la proteína PcpA con la chaoptina y con el receptor de la hormona folículoestimulina humana (Fshr). El porcentaje de similitud entre la chaoptina y la proteína PcpA es de un 67,9% mientras que el de identidad es de un 24,3%. Similares porcentajes de similitud e identidad se establecen entre Fshr y PcpA, siendo éstos de un 66,4% y un 23%, respectivamente. Estas proteínas se caracterizan por contener los motivos anfipáticos

Chaoptina	410	GDRNDPTNLQTLRNMTRMRNMRSLSISRLGSSSVGPED--FKDFGVELEDLQITRAS-LSG	467
PcpA	142	SSITIPSSIKKIQKKGPHGSKAKTIIFDKGSQLEKIEDRAF-DFS-ELEEEEL-PAS-LEY	198
Fshr	26	SNRVFLCQESKVTEIPSDLPNAIELRFVLTKLRVIQKGAFFSGFG-DLEKIEISQNDVLEV	85
Chaoptina	468	IQSHAFFKHVRGLKRLDFSENG-ISSIENDAFHEIGHSLISLKMSHGYSGSALPAEPLRHLT	527
PcpA	199	IGTSAFSSFSQKLKLTFTSSSSKLELISHEAFANLSN-LEKLTLPK--SVKTLGSNLFRLTT	256
Fshr	56	IEADVFSNLPKLHEIRIEKANNNLLYNPEAFQNLN-LQYLLISNT-GIKHL-PDVHKL-H	142
Chaoptina	528	SLQEIDFSNNHISSMSDTSFIH-----FLKNLRLELHDNRIEQVLKGTFGDIHSKLEEI	582
PcpA	257	SLNMLMLRGMIVASVDGVSFQSKTQLIYYPSQKNDES YKTPKETKELASYSFNKNSYLKKL	317
Fshr	143	SLQKVLLDIQDNINIHTIERNSEFVGLSF----ESVILWLNKNGIQEIHNCAFN-GTQLDEL	198
Chaoptina	583	SLRFN--HITSISQHTFFDLEALRKLHLDDNKIDKTIERRAFMNLDELEYLSLRGNKINNLA	641
PcpA	318	EIN-E--GLQKIGTFAFADATKLEEISLPNS-LETIERLAFYGNLELKEILIP-DNVKNFG	373
Fshr	199	NLS-DNNNLEELPNDVFHGAS--GPVILDIS-RTRIHSLSYGLNENLKKL----RARSTYN	251
Chaoptina	642	DESEQNLPK-LEILDMAFNQLPNFNFDYFDQVGTLSNLN-V-NVSHNQIRQLMYNSSWSGR	699
PcpA	374	KHVMNGLPKFLTLSGNNINSLPSFFLS--GVLDLKEIH-IKNKSTEPS--VKKDTFAIPE	429
Fshr	252	LKKLPTLEKLVALMEASLTYP-SHCCAFANWRRQISELHPICNKSILRQE-VDYMTQTRGQ	310

B

LRG	P x x L L x x L - - x - - x L x x L D L S x N - x L x x L
Chaoptin	P x x x F x x L - - x - - x L x x L D L S x N - x L x x I
GPIB	P x G L L x x L - - P - - x L x x L x L S x N - x L T T L
Yop M	x x N x L x x L - - P - - E L x x - - - - - x I x x I
InIA	D I x P L x x L T N L x x L x L x x N - Q I x - -
PcpA	x x x A F x x x - - x - - x L x x L x L x x x - x L x x I
Consenso	* * * * a * * a * * x * * x L x x L x L x x N * x a x x a

Figura 36- Comparación de la secuencia de aminoácidos de PcpA con otras proteínas.

A) Comparación de la secuencia de aminoácidos del dominio N-terminal de la proteína PcpA con la chaoptina de *Drosophila* y con el receptor de la hormona folículoestimulina humana (Fshr). Las barras verticales hacen referencia a los aminoácidos idénticos mientras que los dos puntos hacen referencia a los aminoácidos conservados. B) Se muestra la comparación de las secuencias consenso de algunas proteínas que poseen motivos similares, ricos en leucina (LRRs): la glicoproteína- α_2 humana, la chaoptina de *Drosophila*, la glicoproteína Ib humana (GPIB), la proteína M de la membrana externa de *Yersinia pestis* (Yop M), la internalina (InIA) de *Listeria monocytogenes* y la PcpA de *S. pneumoniae*. Debajo de estas secuencias aparece la secuencia consenso obtenida por Kobe y Deisenhofer (1994) tras la comparación de las LRRs de todas las proteínas con dichos motivos descritas hasta el momento. Los símbolos utilizados: x, aminoácidos no consenso; -, huecos que son necesarios abrir para alinear con la consenso; *, un aminoácido puede estar o no presente en esta posición; a, un aminoácido hidrofóbico, predominantemente leucina, pero puede ser también fenilalanina, isoleucina ó valina.

repetidos ricos en leucina, estableciéndose la homología entre ellas a este nivel. Repeticiones similares a éstas se identificaron por primera vez en una proteína del suero humano con función desconocida, la glicoproteína α_2 rica en leucina (LRG) (Takahashi y cols., 1985). Actualmente están descritas un amplio número de proteínas con diferentes funciones y localizaciones, pertenecientes a una gran variedad de microorganismos, que constituyen una superfamilia de proteínas con motivos ricos en leucina (LRRs) (Para una revisión ver Kobe y Deisenhofer, 1994). Estos motivos destacan por la presencia de una secuencia consenso característica (Figura 36B)

3.5- Caracterización de la unidad de transcripción del gen *pcpA*

Se llevó a cabo el estudio de la transcripción del gen *pcpA* en la cepa R6 de neumococo mediante la extracción de RNA a lo largo del crecimiento y posterior análisis del mismo por *Northern-blot*, utilizando como sonda el fragmento *XbaI-EcoRV* del gen *pcpA* (Fig. 31). El gen parece transcribirse durante la fase exponencial de crecimiento ya que las señales de hibridación sólo aparecen en las muestras extraídas durante esta fase (Fig. 37). El tamaño del transcrito, aproximadamente de 2,3 kb, coincide con el esperado considerando el sitio de iniciación de la transcripción, determinado previamente por *primer extension* (Fig. 33), y la secuencia del posible terminador (Fig. 32). Aparentemente, en estas condiciones el gen *pcpA* se transcribe como un gen monocistrónico. Es posible que en la secuencia palindrómica situada cerca de la región promotora se una alguna proteína responsable de la regulación de la transcripción del gen *pcpA*.

3.6- Obtención de mutantes del gen *pcpA*.

Con la finalidad de conseguir alguna información adicional sobre la funcionalidad de la proteína PcpA, en base tanto a la observación de variaciones en el fenotipo como de alteraciones fisiológicas, se llevó a cabo la obtención de un mutante defectivo en el gen *pcpA* tras la inactivación del mismo por mutagénesis insercional. Para ello fue necesario construir un plásmido que fuera incapaz de replicarse en neumococo, y que contuviera

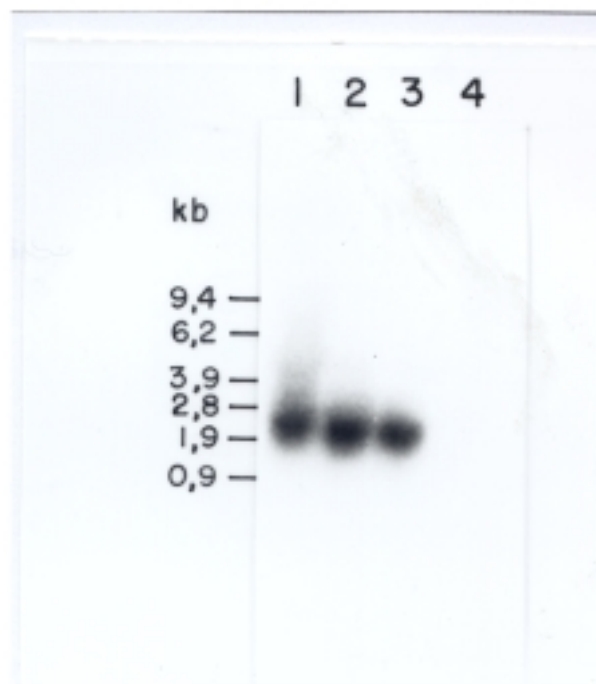


Figura 37- Análisis mediante *Northern-blot* de la unidad de transcripción del gen *PcpA*. Calle 1, RNA extraído a 168 N; Calle 2, RNA extraído a 344 N; Calle 3, RNA extraído a 800 N; Calle 4, RNA extraído a 904 N. En el margen izquierdo de la figura se muestran los tamaños en kb de los marcadores de RNA utilizados.

clonado un fragmento interno del gen *pcpA* unido a un marcador de resistencia que nos permitiera, posteriormente, suponiendo que tuviera lugar un proceso de recombinación homóloga, seleccionar los mutantes. Como vector de clonación se utilizó el plásmido pUCE191, el cual deriva de pUC19 y que lleva clonado el gen de resistencia a la eritromicina procedente del plásmido pFB9 (Barany y cols., 1982). A este plásmido se ligó el fragmento *XbaI-EcoRV* de 772 pb del gen *pcpA*, transformándose posteriormente *E. coli* TG-1 y obteniéndose el plásmido recombinante pEK14 (Fig. 38). Dicho fragmento no contiene ninguna secuencia codificante para los módulos de unión a colina con el fin de evitar recombinaciones inespecíficas en otras regiones del cromosoma donde están presentes genes con secuencias homólogas a dichos motivos. La cepa rugosa M24 de *S. pneumoniae* (García y cols., 1993) se transformó con el plásmido pEK14, el cual debería integrarse por recombinación homóloga dando lugar a la formación de co-integrados entre el plásmido y el genoma, flanqueados por repeticiones directas constituidas por el DNA homólogo utilizado para la integración (Fig. 40).

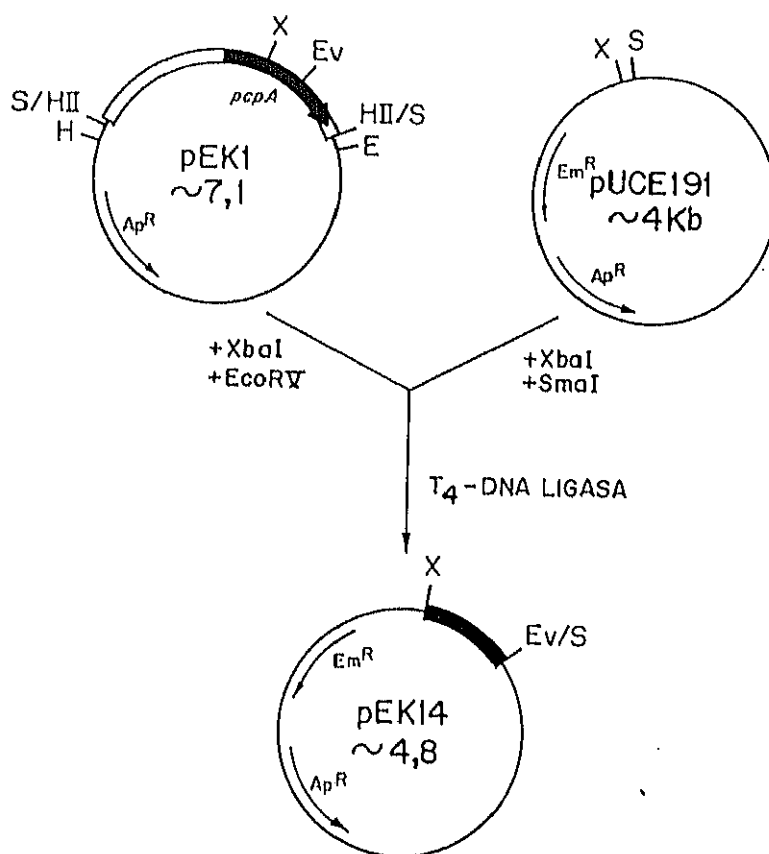


Figura 38- Construcción del plásmido pEK14. La línea gruesa corresponde al inserto de DNA de la cepa de neumococo R6 y la flecha negra al gen *pcgA*. El trazo fino corresponde al DNA plasmídico. Los símbolos Em^R y Ap^R significan resistencia a eritromicina y ampicilina, respectivamente. Dianas de restricción: E, *EcoRI*; H, *HindIII*; S, *SmaI*; H, *HindIII*; X, *XbaI*.

Para demostrar la formación de dichos co-integrados, se llevó a cabo la extracción de DNA de varios de los transformantes seleccionados en medio agar-sangre con lincomicina. Posteriormente se digirieron con *HincII* y *EcoRV* y se analizaron por *Southern-blot*, utilizando como sonda el fragmento *XbaI-EcoRV* del pEK14 (Fig. 38). Todos los transformantes analizados presentaron el mismo perfil de hibridación. Con la digestión *HincII* aparecen tres bandas de hibridación de aproximadamente 5,2 kb, 2,5 kb y 1,8 kb (Fig. 39). En el caso de que sólo se hubiera integrado una copia del plásmido mediante recombinación homóloga, deberían aparecer dos bandas de 5,2 kb y 1,8 kb, en vez de tres. El hecho de que aparezcan tres bandas implica que ha tenido lugar, por un lado la

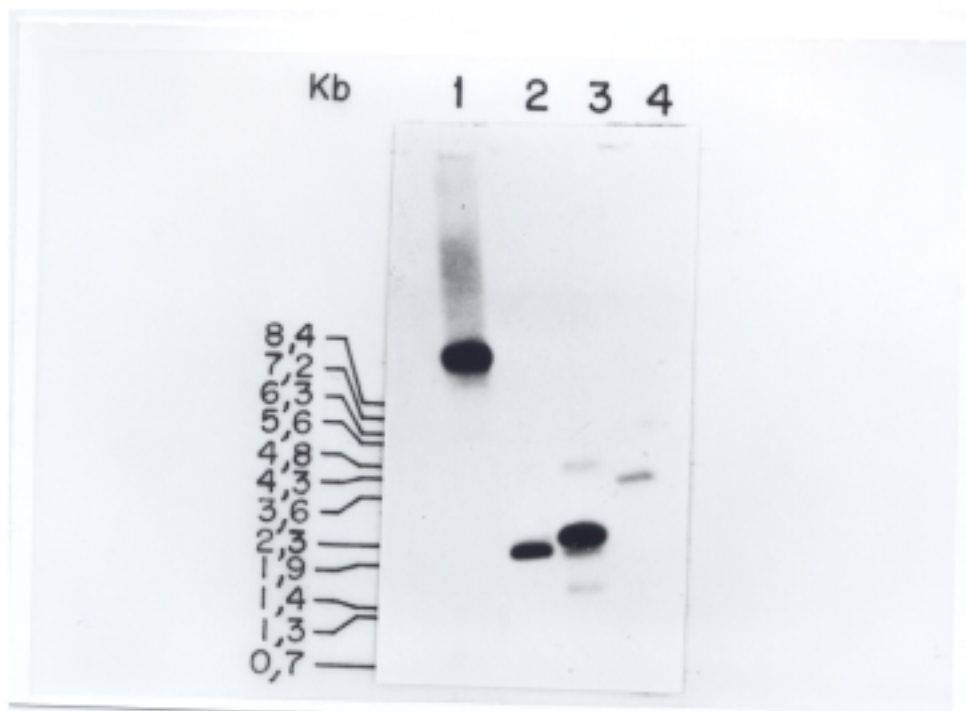


Figura 39- Análisis mediante *Southern-blot* del DNA del mutante K1. Calle 1, DNA del mutante de neumococo K1 digerido con *EcoRV*; Calle 2, DNA de la cepa R6 digerido con *EcoRV*; Calle 3 DNA del mutante de neumococo K1 digerido con *HincII*; Calle 4, DNA de la cepa R6 digerido con *HincII*. En el margen izquierdo de la figura se muestran los tamaños en kb de los marcadores de DNA utilizados.

integración por recombinación homóloga (Fig. 42b) y, por otro lado la amplificación del plásmido integrado debido a la presión selectiva impuesta por el gen de la eritromicina (Mollet y cols., 1993) (Fig. 42c). La intensidad de la banda de 2,5 kb es muy superior a la de las otras dos, lo que implica que tras la integración han tenido lugar varias amplificaciones con la aparición de varias copias de dicho inserto (Figura 42c). Algo similar ocurre con la digestión *EcoRV* de manera que, al digerir el DNA del mutante con dicha enzima, si no se hubiera amplificado el plásmido en el genoma bacteriano, debería aparecer una única banda de un tamaño de 7,1 kb y, sin embargo, aparece una banda de tamaño superior. Un sólo proceso de amplificación daría lugar a una banda de 11,9 kb al digerir con *EcoRV* el DNA cromosómico del mutante (Figura 42c) pero, al igual que en el caso anterior, la banda es muy intensa lo que sugiere que han tenido lugar varias amplificaciones.

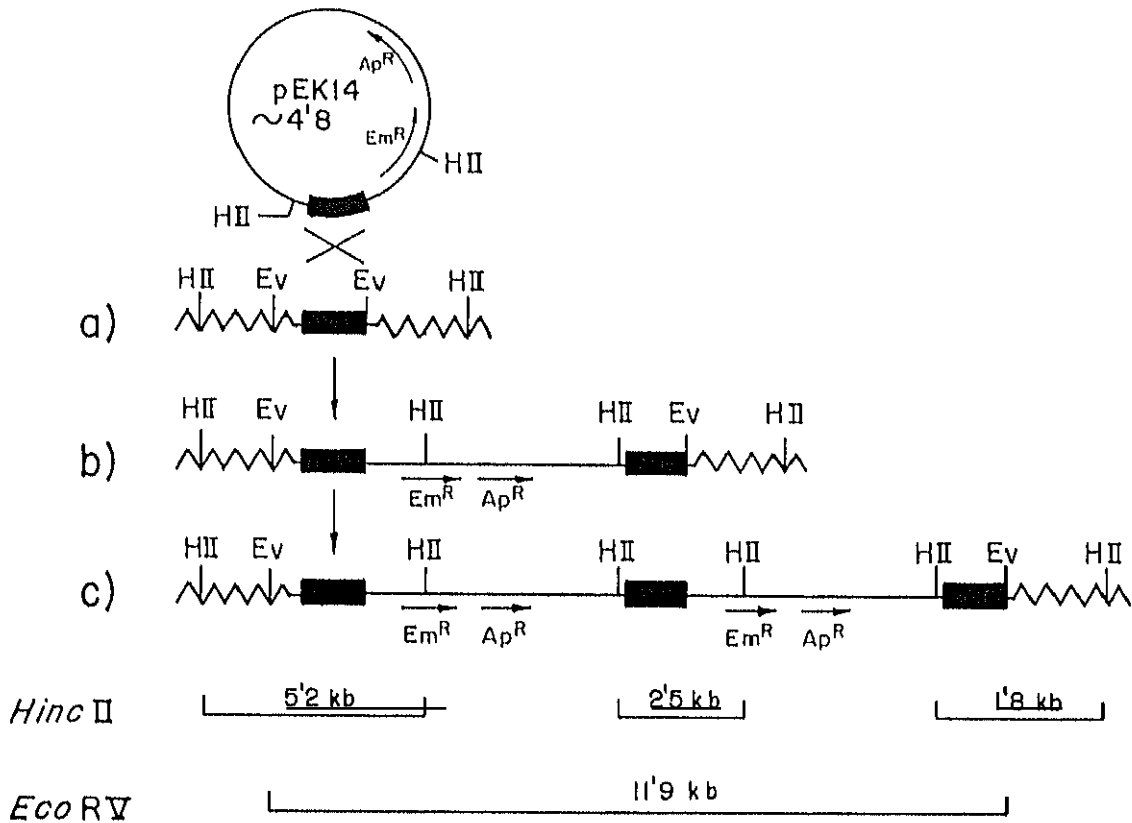


Figura 40- Estrategia de obtención de un mutante por inserción del gen *pcpA*. Dianas de restricción: *HII*, *HincII*; *Ev*, *EcoRV*. a) Proceso de recombinación homóloga entre el fragmento del gen *PcpA* clonado en el plásmido pEK14 y el mismo fragmento presente en el DNA de neumococo. b) Integración del plásmido en el genoma bacteriano tras el proceso de recombinación. c) Amplificación del número de copias del plásmido integrado por recombinaciones intragénicas. La parte inferior corresponde a los fragmentos resultantes de la digestión del último cointegrado con *HincII* y *EcoRV*, respectivamente. Los símbolos Em^R y Ap^R significan resistencia a eritromicina y ampicilina, respectivamente.

Tras la confirmación de que los clones obtenidos eran mutantes de inserción defectivos en el gen *pcpA*, se pudo observar que aparentemente no existían diferencias morfológicas entre la cepa mutante, a la cual hemos denominado K1, y la cepa control M24. Por otro lado, para descartar la posibilidad de que el gen clonado fuera el gen que codifica para la glucosaminidasa, se ensayó la actividad lítica con un extracto de dicho mutante y de la cepa M24 control a 30°C, temperatura a la cual se manifiesta la actividad de esta enzima. El mutante presentaba la misma actividad glucosaminidasa que el control, de lo que se

deduce que el gen inactivado no codifica para esta enzima lítica, a no ser que exista en neumococo otra enzima lítica, aún no descrita, que enmascarase su actividad.

3.7- Clonación del gen *pcpA* en neumococo

Al no poder obtener ninguna información adicional con el mutante K1 sobre la posible función de esta proteína, decidimos clonar el gen *pcpA* en neumococo. Al existir un mayor número de copias del gen en la cepa recombinante, la proteína se sintetizaría en mayor cantidad, lo que podría dar lugar quizás a variaciones fisiológicas que nos permitieran aportar alguna información adicional sobre la funcionalidad de la misma. El plásmido utilizado para la transformación de neumococo fue el pEK15, que presenta un tamaño de 9,6 kb. Para la construcción de dicho plásmido se siguió la estrategia que se muestra en la figura 41, en donde el fragmento *SaI*-*SacI* de 4,4 kb del pEK1 se clonó en el plásmido pLSE3 (Díaz y García, 1990a). La mezcla de ligación se utilizó para transformar *E. coli* C600, seleccionándose los clones con el plásmido recombinante pEK15 mediante hibridación de colonias, utilizando como sonda el fragmento *XbaI*-*EcoRV* del plásmido pEK14 (Fig. 38). Posteriormente se transformó con dicho plásmido la cepa de neumococo M24 y se estudió la curva de crecimiento tanto a 30°C como a 37°C. A 30°C la lisis del cultivo se manifestó de igual manera en las cepas M24, M24 (pEK15) y K1, por lo que descartamos la posibilidad de que el gen *pcpA* codifique para la glucosaminidasa.

3.8- Localización del gen *pcpA* en el genoma de *S. pneumoniae*

El desarrollo reciente del mapa físico de la cepa R6 de *S. pneumoniae* (Gasc y cols., 1991) ha proporcionado una herramienta útil para la localización cromosómica de los diferentes genes. El DNA cromosómico de la cepa R6 de *S. pneumoniae* fue digerido con las enzimas de restricción *SmaI*, *SacII* y *ApaI*, sometiéndose a una electroforesis en campo pulsado (PFGE). El DNA se transfirió a una membrana de nylon y se hibridó con el fragmento *XbaI*-*EcoRV* del plásmido pEK1. En la figura 42 se muestra el perfil de hibridación, localizándose dicho gen en los fragmentos quinto (de la digestión *SmaI*) y

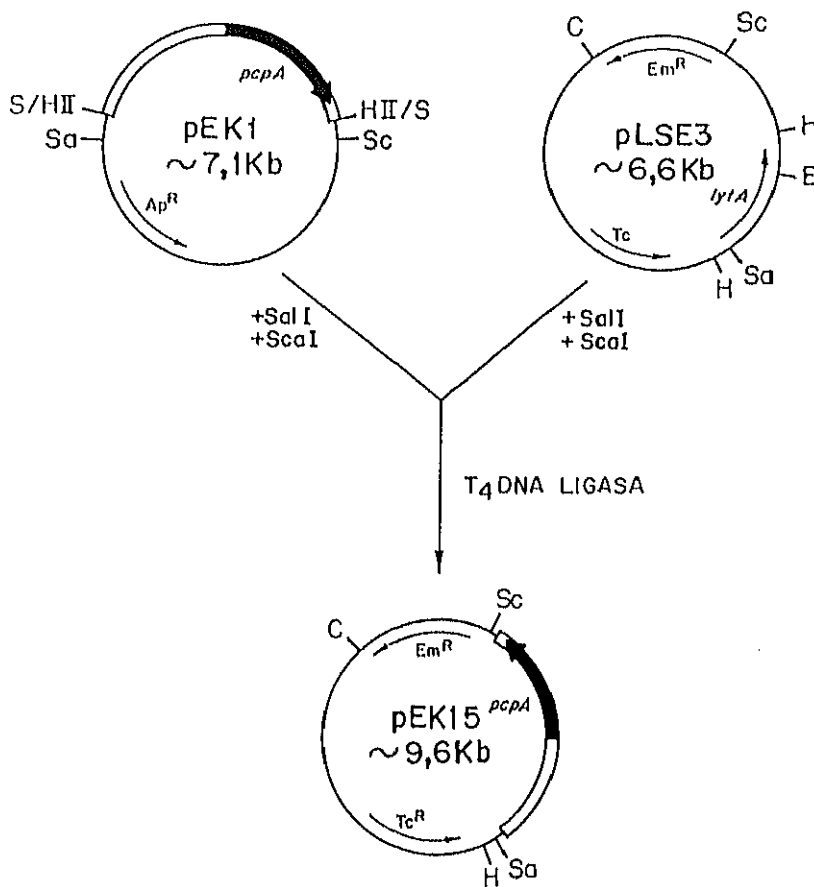


Figura 41- Construcción del plásmido pEK15. Dianas de restricción: E, *EcoRI*; H, *HindIII*; S, *SmaI*; C, *ClaI*; Sa, *SalI*; Sc, *SacI*. La trazo grueso corresponde al inserto de DNA de la cepa de neumococo R6 y la flecha negra al gen *pcgA*. El trazo fino corresponde al DNA plasmídico. Los símbolos Em^R, Ap^R significan resistencia a eritromicina y ampicilina respectivamente.

cuarto (de la digestión *ApaI*) con un tamaño, en ambos casos, de 235 kb. El gen *pcpA* se localiza por tanto, en el mismo fragmento cromosómico que los genes *hexA* y *malM* implicados en procesos de recombinación y en el metabolismo de la maltosa, respectivamente. Esta localización no se corresponde con la del gen *lytA*, una de las enzimas de neumococo con dominio de unión a colina (Fig. 43).

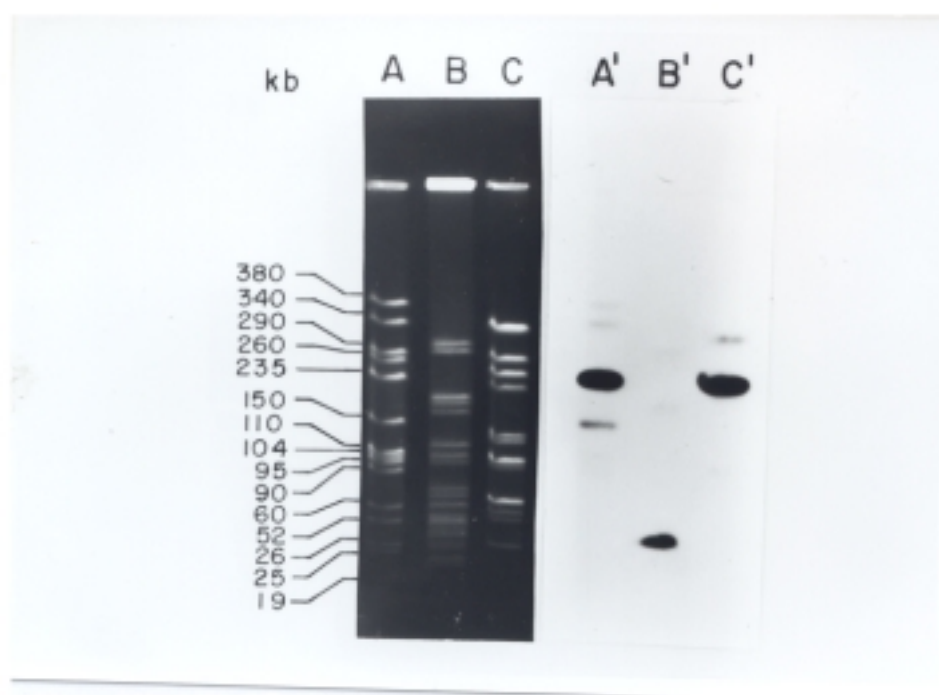


Figura 42- Electroforesis en campo pulsado del DNA de la cepa R6 de *S. pneumoniae*. El panel de la izquierda muestra los fragmentos de restricción separados mediante electroforesis en campo pulsado, teñidos con BrEt y visualizados con radiación UV. El panel de la derecha muestra la autorradiografía del *Southern-blot* de dicho gel tras la hibridación con el fragmento *XbaI-EcoRV* del plásmido pEK1. A y A', digestión del DNA con *SmaI*; B y B' digestión del DNA con *SacII*; C y C' digestión del DNA con *ApaI*. La hibridación y lavados se realizaron a 65°C. A la izquierda se muestran los tamaños de los fragmentos resultantes de la digestión del DNA de la cepa R6 con *SmaI*, expresados en kb.

4- CARACTERIZACIÓN DE UNA NUEVA SECUENCIA DE INSERCIÓN (IS1381) EN *S. pneumoniae*

4.1- Análisis de la secuencia de nucleótidos y aminoácidos: comparación con otras proteínas del banco de datos.

Precediendo al gen *pcpA* (Fig. 31 y 32), se localizan en orientación contraria dos marcos de lectura abiertos (ORF3A y ORF3B) de 570 y 285 nucleótidos, respectivamente. En el gen de la ORF3 B se localiza, a 6 nucleótidos del codón de iniciación, la secuencia de un posible RBS (Fig. 32). No se detectó, sin embargo, la secuencia de un posible RBS precediendo a la ORF3A, lo que hace improbable que esta proteína pueda ser sintetizada (Fig. 32). La ORF3A y la ORF3B codificarían para proteínas de 10,5 y 21,7 kDa con pI de 11,1 y 10,3, respectivamente. Al comparar dichas secuencias con las de la base de datos del EMBL, se comprobó que ambas mostraban una importante semejanza con la posible transposasa presente en la secuencia de inserción IS702 de la cianobacteria *Calothrix* PCC 7601. Al alinear la secuencia de aminoácidos de la ORF3B, con los 134 primeros aminoácidos de la transposasa de IS702, se comprobó que el 32,5% de los residuos eran idénticos y que un 53,4% correspondía a sustituciones conservadas (Fig. 44). Un alineamiento similar entre los aminoácidos 145-235 de la misma proteína, con la secuencia de aminoácidos de la ORF3A, dió un 35,5% de identidad y un 56,6% de similitud (Fig. 44). Si tenemos en cuenta el solapamiento de ambas ORFs mediante un salto en el marco de lectura de un sólo nucleótido (*frameshifting* +1), concretamente en la timina del codón de terminación TAA de la ORF3B (Fig. 32), tendríamos un gen que codificaría para una transposasa de 293 aminoácidos similar a la de *Calothrix* (Fig. 44). Su pI de 10,8 es típico de otras transposasas de secuencias de inserción así como de proteínas que se unen a DNA (Galas y Chandler, 1989). Puesto que la transposasa de la IS702 de *Calothrix* está codificada por un único marco de lectura, la organización en dos genes podría ser un rasgo específico de la IS1381 o podría reflejar simplemente una mutación en la copia analizada. Para comprobar que la mutación no tuvo lugar durante el proceso de clonación, se procedió a secuenciar de nuevo la región de la transposasa, pero esta vez a partir del plásmido pEK1 aislado de un clon diferente. La secuencia resultó ser la misma en ambos casos.

Casi todas las ISs descritas hasta el momento presentan repeticiones terminales invertidas (RI) con un mínimo de 10 nucleótidos, generando además secuencias duplicadas

```

IS1381  51 QRVNRLPAKQNYEATKQLTDARFKRLVGVQRTTFFEEMLAVLKTAYQLKHA 100
          ..| .|||. :||: || | ||: |:|:| | :|.
IS702    1 .....MKIEKAKQLTARKFKRMSGVSRQTFNYMVDVVK.ADEK KKK 40
IS1381 101 KGGRKPKLSLEAFLWPLENMCENIELMNKLRLIFGIHESNLIRRSQWVEV 150
          |.||:|||| :|. :. :. :. :. :. | :|: |||. :|
IS702   41 KPGRRPKLIIEDQVLMVIQYWREYRTYYHIGLDWGLSESAVCRTVYKIEN 90
IS1381 151 TLVQS.....GVTISRTPLSSEDTVMIIDATEVKINRPKKRISEN-SGKK 195
          .|: | ... . :. ||: :. :. :. |. |. |||. . . |||
IS702   91 ILISSRKFSPLPGKKELLKMPSQENLVMDVTESPIERP KKSQKKFFSGKA 140
IS1381 196 KFHAMKAQAIVTSQ.GRIVSLDIADVNYSHDMKLFKMSRRNIEQAGKILAD 244
          |.:|. | :. :. ||: :. :. :. ||: ||| | :. :. ||: ||
IS702  141 GEHTLKTQLVIHQKTSQIICLGHGKGRIHDFRLFKTSGVKFSELLKVIAD 190
IS1381 245 SGYQGLMKIYPQVQTPRKSSKLKPLTAENKAYNHASIKKEARLRTSLPN. 293
          .|||: ||. :|| |.. |.. |.. |.. |.. :. .|||. :. :
IS702  191 KGYQGITKIHKLSETPIKKPKGKKLAKEQKEYN....RELNRLRIVVEHV 236

```

Figura 44- Comparación de la secuencia de aminoácidos de las transposasas de IS1381 de *S. pneumoniae* y de IS702 de la cianobacteria *Calothrix* PCC 7601. La comparación con la ORF3B comprende desde el comienzo hasta la secuencia de aminoácidos marcada en negrita. La comparación con la ORF3A comprende desde la metionina que precede a la secuencia marcada en negrita hasta el final. Si se produjera un salto en el marco de lectura de tipo +1 se produciría un solapamiento de ambas ORFs a través de la secuencia marcada en negrita, siendo el porcentaje de similitud de un 54,1% y el de identidad de un 34%.

del DNA, de 2 a 12 pb, correspondientes a la secuencia de DNA que reconoce previa inserción (repeticiones directas ó RD). En nuestro caso, la IS1381 presenta dos repeticiones directas de 7 nucleótidos, sin embargo, las posibles repeticiones invertidas son muy imperfectas (Fig. 32).

4.2- Frecuencia y distribución de la IS1381 en diferentes cepas de *S. pneumoniae*

Para confirmar la presencia repetida de esta secuencia de inserción, se realizó un *Southern-blot* con el DNA de la cepa de neumococo R6 digerido con varias enzimas de restricción e hibridado con el fragmento *SspI* del plásmido pEK1 que, como se puede ver en la figura 31, contiene la ORF3 completa. El perfil de hibridación demostró la existencia de

al menos seis copias de dicha IS distribuidas por el genoma bacteriano (Fig. 45). Esto viene a demostrar que la *IS1381* ha sufrido transposiciones a diferentes lugares del cromosoma bacteriano.

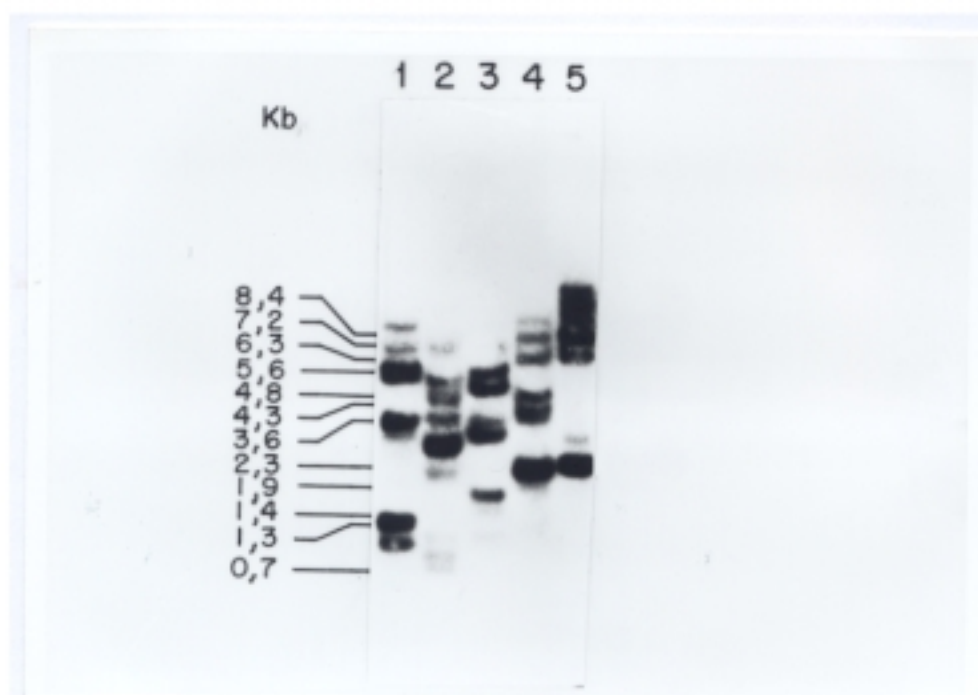


Figura 45- Análisis mediante *Southern-blot* del número de copias de la *IS1381* en la cepa R6 de *S. pneumoniae*. El DNA de la cepa R6 se digirió con *DraI* (calle 1), *EcoRV* (calle 2), *HincII* (calle 3), *HindIII* (calle 4) y *NheI* (calle 5). Se hibridó con el fragmento *SspI* del plásmido pEK1 que contiene el gen de la transposasa de la *IS1381*. La temperatura de hibridación y lavado fue de 65 °C.

Puesto que la cepa R6 presenta múltiples copias de la *IS1381*, decidimos llevar a cabo una exploración inicial de la distribución de este elemento en otras cepas de neumococo. Los DNAs de las cepas de laboratorio R6 y M24 junto con los DNAs de los aislados clínicos S1, D39, 406, S6, S14, S9, S19A y S23, se sometieron a una electroforesis en campo pulsado tras la digestión con *SmaI* y se analizaron mediante *Southern-blot* utilizando la misma sonda que en el caso anterior (Fig. 46). La presencia de la IS se manifiesta no sólo en la cepa R6 sino en todas las demás cepas analizadas. Aunque el perfil de hibridación varía entre unas cepas y otras, todas ellas presentan al menos 3 bandas de hibridación en común con otra cepa. Este perfil es idéntico en el caso de las cepas R6 y

M24 como cabía esperar puesto que la M24 es una estirpe derivada de la R6. Respecto al mínimo número de copias presentes es 6 para todas ellas, excepto 5 para las cepas D39 y S23, y 7 para la S1.

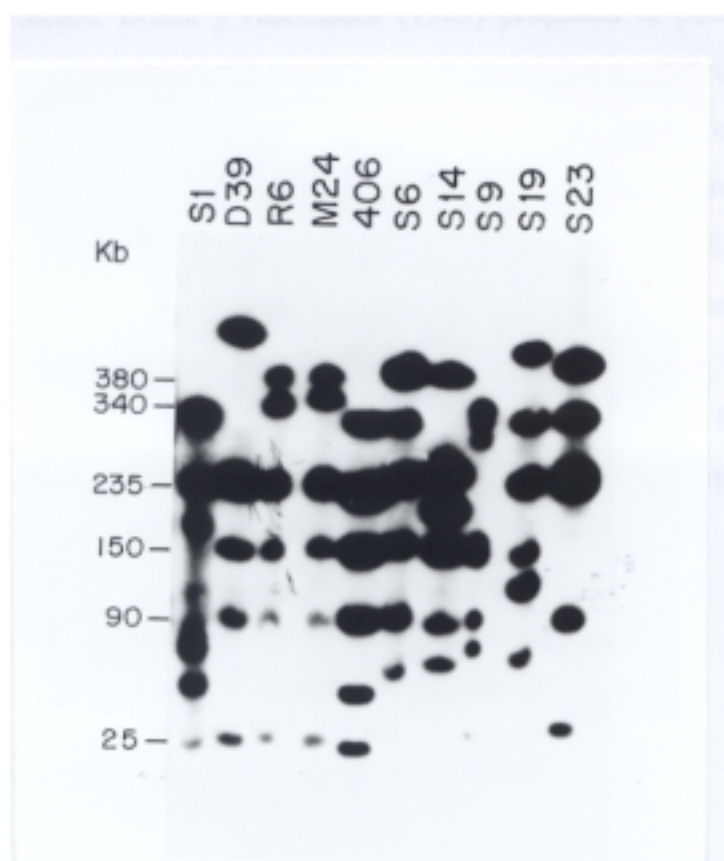


Figura 46- Distribución de la IS1381 en varias cepas de neumococo. En la parte superior de la figura se indican las cepas empleadas. A la izquierda se muestran los marcadores de peso molecular expresados en kb. Se utilizó como sonda en la hibridación el fragmento *SspI* del plásmido pEK1 que contiene el gen de la transposasa de la IS1381. La temperatura de hibridación y la de los lavados fue de 65°C.

5.-CARACTERIZACIÓN MOLECULAR DE LOS GENES *pcpB* y *pcpC*

Hasta la fecha se han secuenciado en neumococo tres proteínas con dominios de unión a colina, LytA (García y cols., 1986b), PspA (Yother y Briles, 1992a) y PcpA descrita anteriormente. Briese y Hakenbeck (1985) proponen la posible existencia de al menos 4 proteínas con dominios de unión a colina en neumococo. En un intento por clonar otros genes que codificasen para proteínas con dominios de unión a colina, se diseñó una nueva estrategia de análisis de genotecas basada en la utilización, como sondas, de oligonucleótidos sintéticos degenerados. Estos oligonucleótidos fueron diseñados en función de los aminoácidos más conservados en los motivos de unión a colina de las proteínas secuenciadas hasta el momento. Para sintetizarlos con la menor degeneración posible, se eligieron los codones utilizados con más frecuencia por los aminoácidos más conservados. De esta manera se diseñaron los oligos Mod1 y Chol1 que se muestran en la figura 47.

Mod1

```

      G  A  M  A  T  G  W
5' -GGTGCTATGGCTACTGGTTGG-3'
      C  A      A  A  A
      A  C      C      C

```

Chol1

```

      W  Y  Y  T/S T/G  N  D  K
5' -ATAATACCA AGA ACC ATTATCTTT-3'
      G  G      GCT G   G  G  C
           C   C
           T   T

```

Figura 47- Secuencia de los oligonucleótidos Mod1 y Chol1. Por encima de la secuencia de nucleótidos se muestra la de aminoácidos en código de una letra.

5.1- Clonación de los genes *pcpB* y *pcpC*

Con anterioridad al análisis de las genotecas, se realizó un *Southern-blot* con el DNA de la cepa R6 de neumococo con el fin de seleccionar las enzimas más adecuadas para la construcción de las mismas (Fig. 48). Aunque el perfil de hibridación fue muy similar, la

intensidad de las bandas no es igual con Mod1 que con Choll. Las enzimas de elección fueron *HincII* (Figura 48, calle 2) y *HindIII* (Fig. 48, calle 3). Con la enzima *HincII* se utilizó toda la digestión para construir la genoteca, sin embargo, con la *HindIII* se purificaron mediante β -agarasa los fragmentos de DNA comprendidos entre 2,3 y 1,2 kb para descartar los genes ya clonados hasta el momento, cuyo tamaño se conocía por su secuencia. A este nivel aparece una banda de hibridación muy gruesa, aunque un análisis detallado permitió distinguir dos bandas muy juntas, una de las cuales podría corresponder, por tamaño, al gen *lytA*. Se construyeron dos genotecas en el vector pUC18 previamente digerido con *SmaI* y *HindIII*, respectivamente, utilizando *E. coli* DH-5 α como cepa receptora.

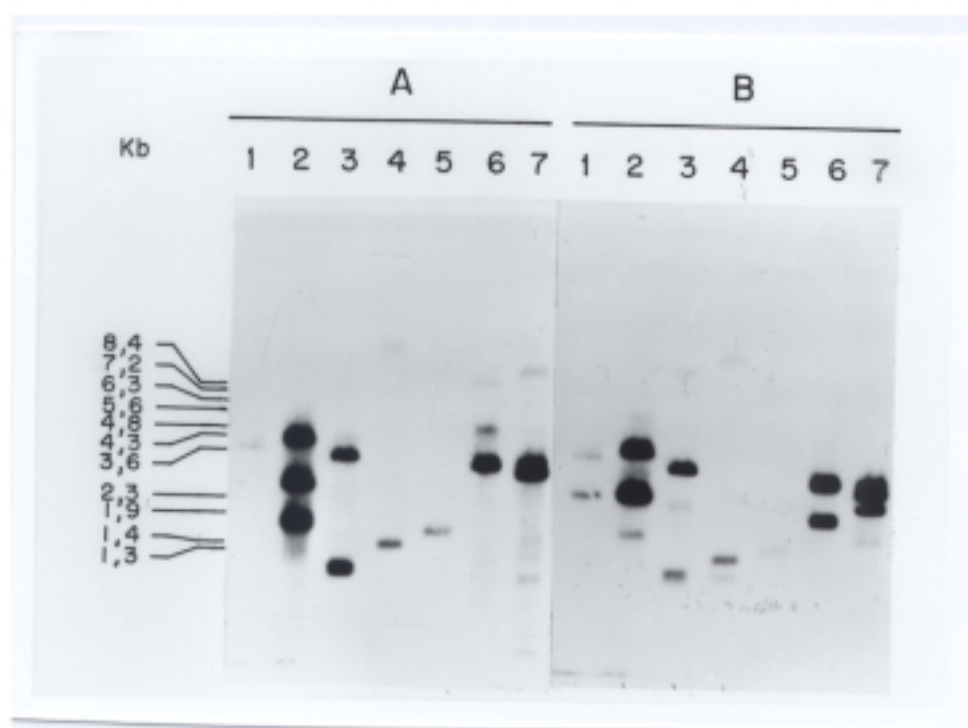


Figura 48- Análisis mediante *Southern-blot* del DNA de la cepa de neumococo R6. A) *Southern-blot* hibridado con el oligonucleótido Choll. B) *Southern-blot* hibridado con el oligonucleótido Mod1. El DNA de la cepa R6 se digirió con las enzimas de restricción *EcoRV* (calle 1), *HincII* (calle 2), *HindIII* (calle 3), *NheI* (calle 4), *PstI* (calle 5), *PvuII* (calle 6) y *SspI* (calle 7). Las hibridaciones y los lavados se realizaron a 40 °C en el caso de Choll y 52°C en el caso de Mod1. Los marcadores de peso molecular se muestran a la izquierda de la figura expresados en kb (λ BstEII).

El análisis de las genotecas se realizó mediante hibridación de colonias, en condiciones no estrictas y utilizando ambos oligonucleótidos como sonda. Se construyeron los mapas de restricción de los plásmidos recombinantes, aislados a partir de los clones positivos obtenidos de ambas genotecas, para descartar aquellos clones que contuvieran las secuencias de los genes ya clonados. De esta manera se aisló el plásmido recombinante pKJ1 de la genoteca *Hind*III, con un inserto de 1,3 kb (Fig. 49). El inserto de dicho plásmido hibridaba con ambos oligonucleótidos y contenía parte de un gen codificante para una nueva proteína con motivos de unión a colina. Este gen recibiría más tarde la denominación *pcpC*.

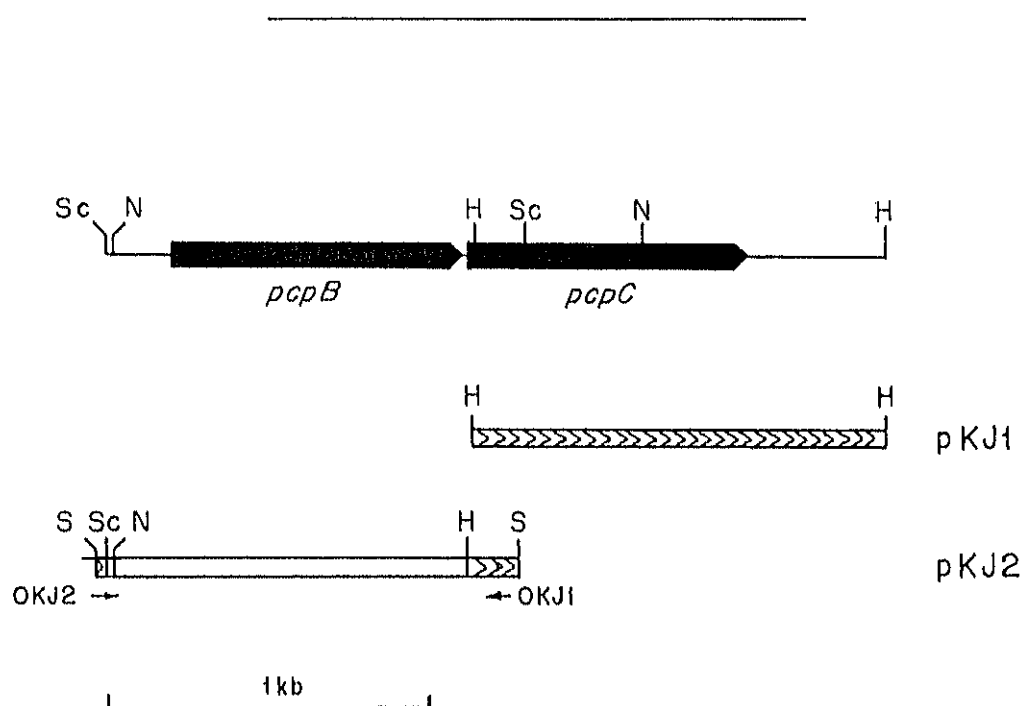


Figura 49- Mapa de restricción resultante del solapamiento de los insertos de los plásmidos pKJ1 y pKJ2. En la parte superior de la figura se muestra el fragmento de restricción *ScdI*-*Hind*III del DNA de la cepa de neumococo R6 resultado del solapamiento de los insertos presentes en los plásmidos pKJ1 y pKJ2 que se muestran en la parte inferior de la figura. OKJ1 (5'-GAGCCCCCGGGGGTAGTAGGTTCTACC-3') y OKJ2 (5'-GAAACCCCCCGGGGAGTGATGATGGG-3') hacen referencia a los oligonucleótidos sintéticos utilizados para la amplificación por IPCR. Dianas de restricción: Sc, *ScdI*; N, *NotI*; H, *Hind*III.

Puesto que la secuencia del gen *pcpC* estaba incompleta, se procedió a la clonación del extremo 5' de dicho gen mediante la técnica del PCR inverso siguiendo una estrategia similar a la que se muestra en la figura 6 (apartado 1.1). La elección de la enzima de restricción se realizó tras el análisis mediante *Southern-blot* del DNA de neumococo digerido con diferentes enzimas de restricción e hibridado con el fragmento *ScaI-PstI* del plásmido pKJ1. La enzima escogida fue *ScaI* ya que el fragmento tenía un tamaño adecuado, tanto para contener el resto de la secuencia del gen que nos faltaba como para ser amplificado por PCR. El DNA de neumococo digerido con dicha enzima y recircularizado se utilizó como molde para la amplificación con los oligonucleótidos OKJ1 y OKJ2 diseñados con una diana *SmaI* para facilitar su clonación (Fig. 49). Las condiciones de PCR utilizadas fueron las mismas que en el caso del gen *cspA* pero con una temperatura de anillamiento de 58°C. El fragmento amplificado de 1,3 kb se clonó en el vector pUC18 digerido con *SmaI*, empleando como cepa receptora *E. coli* DH5 α . Mediante hibridación, utilizando como sonda el fragmento *ScaI-PstI* del plásmido pKJ1, se aisló el plásmido pKJ2 (Fig. 49) que contiene la secuencia del extremo 5' del gen *pcpC* y una secuencia codificante para otra nueva proteína con dominio de unión a colina a la que hemos denominado PopB.

5.3- Análisis de la secuencia de nucleótidos y aminoácidos

En la figura 50 se muestra la secuencia obtenida a partir de los fragmentos clonados en los plásmidos PKJ1 y PKJ2. En ella podemos distinguir dos marcos de lectura abiertos. La primera ORF comprende 825 nucleótidos localizándose, a 11 nucleótidos del codón de iniciación ATG, la secuencia de un posible RBS (GGAG) (Murray y Rabinowitz, 1975). Precediendo al RBS se localizan dos secuencias características de regiones promotoras reconocidas por la subunidad σ -70 de la RNA polimerasa de *E. coli*. A 18 pb del codón de terminación TAG se inicia con otro ATG el segundo marco de lectura abierto, constituido por 775 nucleótidos y precedido a 9 pb por la secuencia de un posible RBS idéntico al anterior. A 134 pb del codón de terminación TAA de esta segunda ORF, aparece una secuencia palindrómica capaz de formar una estructura en bucle con una energía libre teórica de -19 kcal/mol que puede actuar como un terminador de la transcripción de tipo rho-independiente. Los genes de ambas ORFs se denominaron *pcpB* y *pcpC* y codifican para dos proteínas de 274 y 251 aminoácidos con pesos moleculares de 31 y 29 kD y con pI

de 7,2 y 6,1, respectivamente. El extremo N-terminal de la proteína PcpC presenta las características propias de un péptido señal, teniendo una región con residuos cargados que preceden a una región hidrofóbica, seguida de una zona con residuos polares.

```

1   AGTACTTTATTGGCTAGCCAGTTATCGACAGTATCTGCTTTGAGTGTTATTTCTAGTACA
61  GGTGAAGAATATGAGGTAAGTGAGACACTAGAAAAAGGTCCAGAGTCTAATGATTCTTCA
121 TTATCTGAGATTTACCAACGATGGTTAACTACCAAAAGCAATCAGAAGTATTATCG
181 GTAATGATGATTTGATTAAATGATAACTTTACAGTATCCATATTCGACTTCAGCTA
      -35                      -10
241 TGGTTTTATCTAAGTATTATGGAGTAGCCGATGGAATGAATGTAGAAGGGAGGGGAAGTG
1'   RBS                      M N V E G R G S
                        ----> PcpB

301 CGAATTTTATTAAAGATAATGTGTTAATTACAGCGGCTCACAACACTACTACAGACATGACT
9'   A N F I K D N V L I T A A H N Y Y R H D

361 ATGGGAAAGAAGCGGATGATATTTATGTTCTTCCGGCTGTTAGTCCAAGTCAAGAAGTAT
29'  Y G K E A D D I Y V L P A V S P S Q E L

421 TTGGAAAAGATCAAAGTAAAGGAAGTTCGTTATTTGAAGGAATTTAGAAATTTAAATTCTA
49'  F G K I K V K E V R Y L K E F R N L N S

481 AGGATGCAAGGGAATATGACTTGGCTTTATTAATTCTAGAAAAGCCCATTTGGTGCAAAAT
69'  K D A R E Y D L A L L I L E K P I G A K

541 TAGGGACTTTGGGTCTTCTACTAGTCAAAAAAATTTGACAGGAATAACTGTGACTATCA
89'  L G T L G L P T S Q K N L T G I T V T I

601 CAGGCTATCCATCATATAATTTTAAATTCATCAAATGTATACAGATAAAAAACAAGTTT
109' T G Y P S Y N F K I H Q M Y T D K K Q V

661 TAAGTGATGATGGCATGTTCTTGGATTACCAAGTTGATACTTTAGAGGGGTCTAGTGGAT
129' L S D D G M F L D Y Q V D T L E G S S G

721 CTACAGTTTATAGTGCTAGTCACCGTGTAGTAGGAGTGCATACTTTAGGAGATGGAGCTA
149' S T V Y S A S H R V V G V H T L G D G A

781 ATCAAATTAACAGTGCAGTTAAATTAATGAAGCAAATTGCCATTTACTTATTTATTCGG
169' N Q I N S A V K L N E A N C H L L I Y S

841 TTCTTAAAGGTTACTCTCTTGAAGGATGGAAGAAAATAAATGGTAGTTGGTACTATTATA
189' V L K G Y S L E G W K K I N G S W Y Y Y
      ----> P1

901 GACAACATGATAAACAACGGGTTGGCAGGAGATAAATGATACTTGGTATTATTTAGACA
209' R Q H D K Q T G W Q E I N D T W Y Y L D
      ----> P2

961 GTTCCGGTAAGATGCTTACAGATTGGCAAAAAGTAAATGGAACTGGTATTATCTCAATT
229' S S G K M L T D W Q K V N G N W Y Y L N
      ----> P3

1021 CAAATGGAGCAATGGTTACAGGTAGCCAAACTATCGATGGTAAAGTTTATAACTTCGCTT
249' S N G A M V T G S Q T I D G K V Y N F A
      ----> P4

1081 CATCTGGTGAGTGGATTGTATGTTGGAGGATATATAAATGAAGCTTTTCAAAAAAATGA
269' S S G E W I * RBS M K L F K K M 7'
                        ----> PcpC

1141 TGCAAGTCGCACTAGCCACATTTTCTTCGGTTTGCTAGGGACCAGTACAGTATTTGCAG
      M Q V A L A T F F F G L L G T S T V F A 27'

```

1201	ATGATTCTGAAGGATGGCAGTTTGTCCAAGAAAATGGTAGAACCTACTACAAAAGGGGG D D S E G W Q F V Q E N G R T Y Y K K G 47'
1261	CTCTAAAAGAAACCTACTGGAGAGTGATAGATGGGAAGTACTATTATTTTGATCCTTTAT A L K E T Y W R V I D G K Y Y Y F D P L 67'
1321	CCGGAGAGATGGTTGTCTGGCTGGCAATATATACCTGCTCCACACAAGGGGGTTACGATCG S G E M V V G W Q Y I P A P H K G V T I 87'
1381	GTCCCTCTCCAAGAATAGAGATTGCTCTTAGACCAGATTGGTTTATTTTGGTCAAGATG G P S P R I E I A L R P D W F Y F G Q D 107'
1441	GTGTCTTACAAGAATTTGTGGCAAGCAAGTTTGTAGAAGCAAAAAGTCTACGAATACCA G V L Q E F V G K Q V L E A K T A T N T 127'
1501	ACAAACATCATGGGAAGAATATGATAGCCAAGCAGAGAACGAGTCTATTATTTTGAAG N K H H G E E Y D S Q A E K R V Y Y F E 147'
1561	ATCAGCGTAGTTATCATACTTTAAAAACTGGTTGGATTTATGAAGAGGGTTATTGGTATT D Q R S Y H T L K T G W I Y E E G Y W Y 167'
1621	ATTTACAGAAGGATGGTGGCTTTGATTCTCGCATCAACAGATTGACGGTTGGAGAGCTAG Y L Q K D G G F D S R I N R L T V G E L 187'
1681	CACGTGGTTGGGTAAAGGATTACCCTCTTACGTATGATGAAGAGAAGCTAAAAGCAGCTC A R G W V K D Y P L T Y D E E K L K A A 207'
1741	CATGGTACTATCTAGATCCAGCAACTGGCTGGCAAAACCTTGGGAACAAATGGTACTATC P W Y Y L D P A T G W Q N L G N K W Y Y 227' ----> P1
1801	TCCGTTTCATCAGGAGCTATGGCAACTGGTTGGTATCAGGAAGGTTCGACTTGGTACTATC L R S S G A M A T G W Y Q E G S T W Y Y 247' ----> P2
1861	TAAATGCAAGTAATGGAGATATGAAAACAGGCTGGTTCCAAGTCAATGGTAACTGGTACT L N A S N G D M K T G W F Q V N G N W Y 267' ----> P3
1921	ATGCCTATGATTTCAGGTGCTTTAGCTGTTAATACCACAGTAGGTGGTTACTACTTAAACT Y A Y D S G A L A V N T T V G G Y Y L N 287'
1981	ATAATGGTGAATGGGTAAAGTAATGAAGGCTAATTGTAAACTGTGATCCATACTTAACTT Y N G E W V K * 294'
2041	GTATAATAGGTGGATAAAAGTCTTCACAATTCAAAAAACGCATAGTATCAAGGGTTTTTC
2101	TGTACTGGCCCCCAAAGTTAGACAAATAATTTATCCAAGGATTTAGTTCTGTACTGCT
2161	ACAGACTAAGTCCTTTACTTTTACCTTAATCTCTTGTGTTGTAATAATCAATATAGT
2221	CTATAATGGCTTGGTTCCAATTGATTAAGTGATTTAAATGTTTTCTCATAGCCATAAAACA
2281	TTTCCGATTTTCGATAATCCCAAAGAAGGACTCCACCATAGTGTTCCTCTGTGAATGCCTG
2341	TTCCAAGATTGTTTGTACTTGTTCCTAGTTGGGTGGAAGTTGAAAGATATAGGCGATAATT
2401	TCGCTGTTAAAGTCCATCTAAAAGTGGTGATAAGTAAAGCTT

Figura 50- Secuencia de nucleótidos de los insertos solapantes de los plásmidos pKJ1 y pKJ2. Se muestra la secuencia de nucleótidos de la hebra de DNA codificante y la secuencia de aminoácidos en código de una letra. Los nucleótidos subrayados corresponden a los sitios de unión al ribosoma (RBS), a la posible región promotora responsable de la transcripción y al posible terminador de la transcripción. P1, P2, P3 y P4 hacen referencia a los motivos de unión a colina presentes en estas proteínas. En negrita y subrayado se marcan los principales aminoácidos que constituyen el centro catalítico de la posible proteasa PepB.

La comparación de dichas proteínas con las proteínas de la base de datos del EMBL reveló una alta similitud con los dominios de unión a colina de las enzimas líticas de *S. pneumoniae* y sus bacteriófagos, con la proteína PspA, con las toxinas de *C. difficile* y con las GTFs y GBPs de *Streptococcus* (Fig. 57, Pag. 162). El dominio C-terminal de la proteína PcpB presenta cuatro motivos de unión a colina, si bien el último está algo degenerado (Fig. 51A). Aunque la proteína PcpC presenta en el extremo C-terminal tres

A

I	197	GWKKINGSWYYYRQH-DKQT
II	216	GWQEINDTWYYLDSSGKMLT
III	236	DWQKVNGNWYYLNSNGAMVT
IV	256	GSQTIDGKVYNFASSGEWI

B

I	32	GW-QFVQENGRT-YYKK---GALKET
II	53	YWRVID---GK-YYYFDPLSGEMVV-
III	74	GW-QYI--PAPHK---GVTIGPSPRIEIALRP
IV	100	DW-----FYFGQ-DGVLQEFV
V	114	GK-QVLEAKTATNTNKH--GEEYD
VI	137	S--QA-E-KRV--YYFEDQRSYHTLKT
VII	158	GWIIYEE---GY-WYYLQK-DG----
VIII	174	GFDSRINRL-----TV--GELAR
IX	190	GWVKDYPLTY-DEEKLKAAP-----
X	217	-----WYY---D---PA-T
XI	217	GW-QNL---GNKWYYLRSS-GAMA-T
XII	237	GWYQE---GSTWYYLNASNGDMK-T
XIII	257	GWFQVN---GN-WYYAYDS-GALAV
XV	276	----NTTV GG--YYLNY-NGEWVK

Figura 51- Motivos repetidos presentes en las proteínas PcpB y PcpC. A) motivos de unión a colina presentes en la proteína PcpB. B) Motivos presentes en la proteína PcpC en donde los cuatro últimos guardan la estructura de los motivos de unión a colina mientras que los restantes parecen motivos degenerados. En negrita se hace referencia a los aminoácidos más conservados en los motivos de unión a colina y que están presentes en los motivos degenerados de la proteína PcpC. Los guiones hacen referencia a los espacios que han sido necesario abrir para alinear los aminoácidos de los motivos degenerados con las secuencias consenso de los motivos de unión a colina.

motivos de unión a colina y una cola C-terminal reminiscente de un cuarto motivo, toda la proteína parece estructurarse en *pseudomotivos* resultado de motivos degenerados con menos de 20 aminoácidos de longitud. En estos *pseudomotivos* se encuentran bastante conservadas las glicinas y los aminoácidos aromáticos (Fig. 51B)

La comparación de la región N-terminal de la proteína PcpB con las proteínas de la base de datos del EMBL dió una similitud de un 41% y una identidad del 24% con el precursor de una glutamil endopeptidasa de la cepa de *B. licheniformis* ATCC 14580 y un 50% de similitud y 26% de identidad con una posible serín-proteasa de *Enterococcus faecalis* (Fig. 52).

Blase	101	DRTRVTNTTAYPYRAIVHISSSIGSCTGWMIGPKTVATAGHCIYDTSSGS	150
		: ...: : . . .	
PcpB	1	MNVEGRGSANFIKDNVLITAAHNYRHDYGK	31
		. : .: : . .: :	
SprE	68	DTTEAPFASIGRIISPASKPGYISLGTGFVVGNTIVTNNHVAESFKNAK	117
Blase	151	FAGTATVSPGRNGTSYPYGSVKSTRYFIPSGWRSNGT...NYDY	191
		.: .: ... : .: . . .: . . : .	
PcpB	32	EADDIYVLPVAVSPSQELFGIKVKEVRYLKEFRNLNSKDAREYDL	76
		. :...: : : : . . :	
SprE	118	VLNPNAKDDAWFYPRDGSATPFGKFKVIDVAF.....SPNADI	156
Blase	192	GAIELSE....PIGNTVGYFGYSYTTSSLVGTTVTISGYPGDKTAGTQ.W	236
		: : : . : :	
PcpB	77	ALLILEK....PIGAKLGLTLGLPTSQKNLTGITVTITGYPSYNFKIHQMY	122
		: : : .: .: .: : . : . . :	
SprE	157	AVVTVGKQNDRPDGPPELGEILTPEVLKKFESSDTHVTISGYPGKHNHTQW	206
Blase	237	QHSGPIAIS.....ETYKLQYAMDTYGGQSGSPVFEEQSSSRNTNCSGPCSL	281
		. . : . . : .: : : .:	
PcpB	123	TDDKKQVLSD.....DGMFLDYQVDLTLEGSSGSTVYSASHRVVGVHTLGDG	167
		. . : . . : . : .: : .: . :	
SprE	207	SHENDLETSNFTDLENPLLEFYDIDTTGGQSGSPIYNAQFEVVGVSNGGI	256
Blase	282	AVHTNGVYGGSSYNRGTRITKEVED.NLTNWKNSAQ	316
		:.: : : . .: . .	
PcpB	168	ANQINSAVKLNEANCHLLIYSVLKGYSLEGWKKIN	202
		. : : .: : : . . : : .	
SprE	257	KQTGNHGQRLNEVNYNFIVNRV...NEEENKRLSAVPAA	292

Figura 52- Comparación de la proteína PcpB con las de la base de datos del EMBL.

La comparación se ha hecho tomando como referencia la proteína PcpB. Se compara la secuencia de aminoácidos del dominio N-terminal de la proteína PcpB (residuos 1 al 316) con la del precursor de una glutamil endopeptidasa de *B. licheniformis* ATCC 1458 denominada (Blase)(residuos 101 al 316) y con la secuencia de aminoácidos (residuos 68 al 292) de la posible serín-proteasa producto del gen *sprE* de *E. faecalis*, que comprende desde el aminoácido 68 hasta el último aminoácido de dicha proteína. Las barras verticales corresponden a los aminoácidos idénticos y los puntos a los conservados.

5.3- Localización de los genes *pcpB* y *pcpC* en el genoma de *S. pneumoniae*

Para la localización de los genes *pcpB* y *pcpC* se utilizó como sonda radiactiva el inserto *Hind*III del plásmido pKJ1 (Fig. 48). La membrana que se hibridó fue la misma que se utilizó en la localización cromosómica del gen *pcpA* (apartado 3.8), tras su deshibridación. Los genes se localizan en el tercer fragmento *Sma*I y en el decimotercero *Apa*I (Fig. 53). En este lugar se localizan los genes que intervienen en la síntesis de la cápsula en neumococo (Arrecubieta y cols., 1994) (Fig. 43, pag. 127). En el fragmento *Sma*I contiguo se localiza el gen *lytA* que como ya se ha comentado codifica para una proteína con dominio de unión a colina.

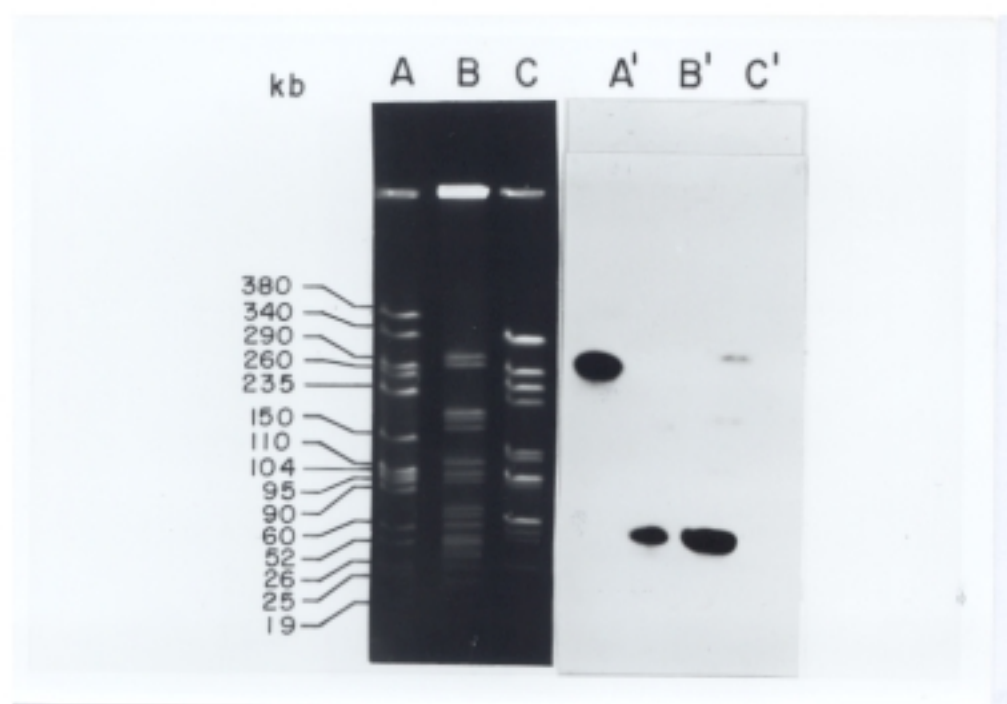


Figura 53- Electroforesis en campo pulsado del DNA de la cepa R6 de *S. pneumoniae*.

El panel de la izquierda muestra los fragmentos de restricción separados mediante electroforesis en campo pulsado, teñidos con BrET y visualizados con radiación UV. El panel de la derecha muestra la autorradiografía del *Southern-blot* tras la hibridación con el fragmento *Hind*III del plásmido pKJ1. A y A', digestión del DNA con *Sma*I; B y B', digestión del DNA con *Sac*II; C y C', digestión del DNA con *Apa*I. La hibridación y lavados se realizaron a 65°C. A la izquierda se muestran los tamaños de los fragmentos resultantes de la digestión del DNA de la cepa R6 con *Sma*I, expresados en kb.

5.4- Análisis del número de genes codificantes para proteínas con dominio de unión a colina en *S. pneumoniae*

Se había visto que al hibridar el DNA de *S. pneumoniae* digerido con *HincII*, con los oligonucleótidos sintéticos Mod1 y Chol1, aparecían dos bandas más intensas y dos muy tenues con el primero y tres muy intensas con el segundo (Fig. 48). Para intentar determinar qué bandas correspondían a los genes ya clonados, se hibridó el DNA de la cepa R6 de neumococo digerido con *HincII* con los genes *lytA*, *pspA* y *pcpC* y con los oligos Mod1 y Chol1 tras la realización de un *Southern-blot* (Fig. 54).

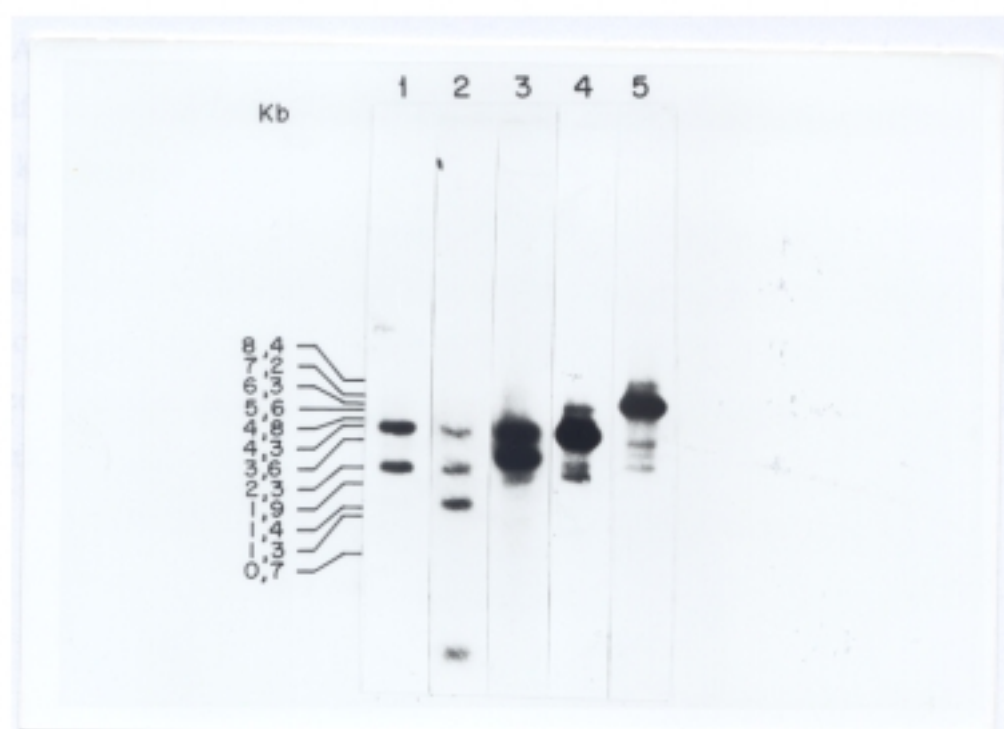


Figura 54- Análisis mediante *Southern-blot* del DNA de la cepa R6 de *S. pneumoniae*. El DNA se digirió con *HincII* y se hibridó en diferentes condiciones con los oligonucleótidos sintéticos Mod 1 (calle 1) y Chol 1 (calle 2), con el gen *lytA* (calle 3), con un fragmento del dominio de unión a colina del gen *pspA* (calle 4) y con el fragmento *HindIII* contenido en el gen *pcpC*. Las condiciones de hibridación y lavados fueron de 50°C cuando se utilizó el oligonucleótido Mod1, de 40°C cuando se utilizó Chol1, y de 65°C en los tres casos restantes.

Se puede comprobar, que al hibridar con el gen *LytA* aparecen dos bandas correspondientes a los dos fragmentos *HincII* que abarcan el gen completo, presentando

una de ellas el mismo tamaño que la banda de 4,4 kb que hibrida con los oligonucleótidos Chol1 y Mod 1 y otra un tamaño un poco más pequeño (Fig. 54, calle 3) En la cepa R6 de neumococo no se sabe cuál de las dos contiene el dominio de unión a colina pero en la cepa DW 3.8 se sabe que está en un fragmento de aproximadamente 4,3 kb. Al hibridar con un fragmento de DNA codificante para el dominio de unión a colina de la proteína PspA, aparece una banda de hibridación también de un tamaño de 4,4 kb (Fig. 54, calle 4) mientras que con el gen PcpC aparece una banda de aproximadamente 5,6 kb (Fig. 54, calle 5). La hibridación con el fragmento *PvuII-SspI* que codifica para el dominio de unión a colina de la proteína CspA de *C. acetobutylicum* NCIB 8052 daba las dos mismas bandas de hibridación intensas que aparecen con el oligonucleótido Mod1, correspondiendo la de 4,4 kb al gen *pcpA* (Fig. 29, calle 4). Por tanto, los genes de neumococo clonados hasta ahora que codifican para proteínas con dominios de unión a colina se corresponderían con la banda de 4,4 kb que aparece al hibridar con ambos oligos y con la banda de 5,6 kb que aparece al hibridar con Mod1. Esto significa que las otras tres bandas de hibridación que aparecen por debajo de la banda de 4,4 kb (Fig. 54, calle 2) codificarían si no para los dominios de unión a colina de otras nuevas proteínas, sí al menos para una sola, que podría ser la glucosaminidasa tan repetidamente buscada. Podemos decir por tanto, que en neumococo existen al menos cinco proteínas con dominios de unión a colina.

IV- DISCUSIÓN

1- CARACTERIZACIÓN MOLECULAR DE LAS PROTEÍNAS Csp DE *C. acetobutylicum* NCIB 8052

Como hemos demostrado con el presente trabajo, la presencia de genes codificantes para proteínas con dominios de unión a colina no es algo restringido a *S. pneumoniae* y sus bacteriófagos, sino que es un hecho evidente asimismo en *C. acetobutylicum* NCIB 8052. Aunque ambos microorganismos ocupan nichos ecológicos muy diferentes, siendo el primero un patógeno humano y el segundo un microorganismo del suelo de interés industrial, ambos poseen en común la peculiaridad de presentar colina en los ácidos teicoicos de pared. La colina parece ser por tanto el elemento que ha ejercido una fuerte presión selectiva permitiendo la conservación de las regiones C-terminales de estas proteínas capaces de reconocer dicho ligando, siendo extrapolables estos resultados a otros microorganismos que contienen colina en las envueltas celulares. Aunque no se han clonado los genes que codifican para dichas proteínas en estos otros microorganismos, en los casos de las cepas de *C. acetobutylicum* NI-4 y NI-504 se ha comprobado que presentan fragmentos de DNA que hibridan con el fragmento de DNA que codifica para el dominio de unión a colina de la proteína CspA (Fig. 14). Por otro lado, aunque se había propuesto que las cepas NCIB 8052 y NI-4 eran idénticas (Wilkinson y cols., 1993), nuestros resultados indican claramente que son diferentes (Fig. 14). Esta discrepancia podría tal vez ser debida a la utilización de cepas procedentes de orígenes diferentes.

En este trabajo se describe la presencia de cuatro genes (*csp*) que codifican para proteínas con dominios de unión a colina y que presentan la peculiaridad de poseer una elevada similitud entre ellos (Fig. 25). La comparación de las regiones no codificantes de estos genes revela, sin embargo, grandes diferencias, lo que sugiere que han divergido más rápidamente que las regiones codificantes o alternatively que las duplicaciones génicas han tenido lugar en zonas próximas a los codones de iniciación y de terminación. Esta observación abre la posibilidad de analizar la evolución de los genes *csp* y particularmente de las regiones que codifican para los dominios C-terminales. Este dominio está construido en base a un número variable de repeticiones, de 4 a 6, homólogas a las previamente descritas en las enzimas líticas de pared dependientes de colina (6 repeticiones) y en la proteína PspA de *S. pneumoniae* (10 repeticiones) (García y cols., 1988b; Yother y Briles, 1992a). Un análisis más detallado de las regiones 3' no codificantes de los genes *cspA* y

cspC, muestra que los 43 primeros nucleótidos son casi idénticos, sugiriendo que estos genes se han duplicado recientemente. Puesto que la proteína CspC contiene 5 repeticiones y la CspA sólo 4, parece que la delección de la última repetición se ha producido por un proceso interno que elimina la secuencia que codifica para la quinta repetición preservando la región 3' no codificante así como la cola del extremo C-terminal. La presencia de una cola C-terminal, producida por una degeneración aparente de la última repetición, se ha observado en los dominios C-terminales de las otras proteínas descritas con dominios de unión a colina. Curiosamente, las repeticiones que se localizan en la misma posición en las proteínas Csp presentan una mayor similitud entre ellas que si se comparan con las repeticiones presentes en cada proteína. Estas observaciones sugieren que los dominios C-terminales proceden de una secuencia ancestral común aunque también parecen indicar que las colas presentes en las proteínas CspA, CspC y CspD pueden ser el resultado de pequeñas delecciones producidas en la región que codifica para el último motivo de la proteína CspB. Variaciones en el número y composición de unidades repetidas se han observado también en las proteínas M de *Streptococcus* (Hollinhead y cols., 1987) y en la proteína extracelular de *S. suis* tipo 2 (Smith y cols., 1993).

Hay que destacar que no es frecuente la presencia de isogenes en un mismo microorganismo y, en el caso de los genes *csp*, se sabe que no constituyen un operón aunque al menos tres de ellos mapeen relativamente próximos en el cromosoma bacteriano (Fig. 27). En *C. acetobutylicum* ATCC 824 se ha descrito también la presencia de dos genes *bdhA* y *bdhB*, que codifican para dos butanol deshidrogenasas homólogas que dependen de NADH y NADPH respectivamente. Ambos genes se disponen uno a continuación del otro, sin embargo, tampoco forman un operón (Walter y cols., 1992). En *E. coli* son pocos los genes estructurales descritos que sean homólogos entre sí y que se localicen en distintos lugares del cromosoma, entre los que se encuentran *argI/argF* (Van Vliet y cols., 1984), *tufA/tufB* (Furano, 1977) y *gadA/gadB* (Smith y cols., 1992). Sin embargo, mientras que los genes *arg* han divergido hasta el punto que las enzimas codificadas por los mismos son bioquímicamente diferenciables, en el caso de los genes *tuf* y *gad* son prácticamente idénticas. En estos genes las diferencias nucleotídicas tienden a encontrarse en los extremos 5' y 3' al igual que ocurre con los genes *csp*. Puesto que los productos de estos genes son indistinguibles funcional y estructuralmente, se sugiere que la presencia de genes duplicados supone una reserva de emergencia para aquellas situaciones en las que se requiere una

fuerte demanda de dichas proteínas. Otro caso interesante es el de las glucosiltransferasas de *S. mutans* GS5 (von Eichel-Streiber, 1992b). Los genes *gtfB* y *gtfC* que codifican para las enzimas GTF-I y GTF-SI constituyen un operón, mientras que la *gtfD* que codifica para la glucosil transferasa GTF-S, se localiza en un lugar diferente del cromosoma (von Eichel-Streiber, 1992b).

El análisis transcripcional realizado en este trabajo demuestra que sólo los genes *cspA* y *cspC* se expresan como unidades transcripcionales independientes durante la fase exponencial de crecimiento (Fig. 26). Esto se corresponde adecuadamente con la presencia de las proteínas codificadas por dichos genes en el sobrenadante del medio de cultivo, constituyendo las proteínas mayoritarias del mismo. Ambas se caracterizan por ser glicoproteínas habiendo sido detectada la presencia de ramnosa al menos en la proteína CspA (Fig. 13). Una característica de estas proteínas es que la movilidad de las mismas en SDS-PAGE, es mucho menor que la que cabría esperar del M_r deducido de la secuencia de aminoácidos (Fig. 10). Esta diferencia no puede ser debida exclusivamente a la presencia de azúcares puesto que la expresión de la proteína CspA en *E. coli* da lugar a una proteína con un peso molecular que, aunque inferior al de la producida en *Clostridium*, es superior al deducido de la secuencia de aminoácidos (Fig. 12B). Casos similares se han descrito en otras proteínas (Smith y cols., 1993) y, en el caso de las proteínas Csp, podría ser atribuible a la elevada carga de las mismas, dado el alto contenido en aminoácidos ácidos. La ausencia de expresión del gen *cspD* y los bajos niveles de expresión del gen *cspB* se corresponden con la ausencia de posibles secuencias promotoras -10 y -35 delante del codón de iniciación de la traducción. Sólo un limitado número de sitios de iniciación de la transcripción se ha determinado mediante mapeo del mRNA en *Clostridium* aunque, en los casos descritos, las cajas -10 y -35 se corresponden con los elementos consenso reconocidos por la RNA polimerasa de *E. coli* y *B. subtilis* (Young y cols., 1989). A pesar de esto, no podemos descartar la posibilidad de que puedan utilizar factores sigma alternativos y/o reguladores transcripcionales que pudieran ser inducidos en diferentes condiciones de cultivo. Una situación similar se observa en el caso de los dos genes que codifican para la proteína de pared de *S. dysgalactiae*, responsable de la unión a la fibronectina, donde sólo uno de los genes se transcribió bajo las condiciones empleadas (Lingren y cols., 1993).

A pesar de que la comparación del dominio N-terminal de estas proteínas con las presentes en el banco de datos del EMBL no proporcionó ninguna pista sobre la posible

funcionalidad de las mismas, si podemos suponer que estas proteínas deben interaccionar con la envuelta celular a través del dominio de unión a colina. Este dominio es completamente funcional dada su unión específica a la DEAE-celulosa o colina-sefarosa (García y cols., 1988b). Además, se ha demostrado mediante un análisis por *Western-blot*, que una fracción de estas proteínas permanece asociada a la envuelta celular (Fig. 10). Por ello las proteínas Csp pueden incluirse dentro de la familia de proteínas asociadas a la pared (WAPs) (Foster, 1993). La mayor parte de las WAPs descritas hasta el momento se encuentran estructuradas en dos dominios, un dominio proporciona el anclaje a la pared celular mientras que el otro es el responsable de las propiedades catalíticas o receptoras. Imitando a la estructura polimérica de la pared, estas proteínas han desarrollado dominios de unión basados en la presencia de motivos repetidos (Wren, 1991). Las GTFs representan un clásico ejemplo de WAPs con una actividad catalítica (Saxena y cols., 1995), mientras que WAPs tales como las proteínas M (Hollinghead y cols., 1987; Hollinhead y cols., 1993), las proteínas de unión a fibronectina (Lingren y cols., 1993; Talay y cols., 1994), las proteínas de unión a glucanos (Banas y cols., 1990), la proteína PspA de neumococo (Yother y cols., 1992a) o el factor proteico extracelular (EF) de *S. suis* (Smith y cols., 1993) se encuentran implicados en procesos de adhesión y/o patogenicidad. La heterogeneidad estructural observada en algunas de estas WAPs parece ser el resultado de procesos de recombinación junto con duplicaciones o deleciones, especulándose sobre la posibilidad de que éste sea un mecanismo que permita a estos microorganismos escapar del sistema inmune del huésped (Hollinghead y cols., 1987; Smith y cols., 1992). Como ocurre con otras WAPs, la función de las proteínas Csp es todavía oscura pero la abundancia de genes repetidos y su presencia en el medio extracelular sugieren que estas proteínas deben desempeñar un importante papel en la supervivencia de este microorganismo.

Se ha demostrado recientemente que el dominio de unión a colina de la proteína de superficie de neumococo, PspA, permite a esta proteína permanecer anclada a la superficie celular bacteriana. Además, la delección parcial de este dominio así como las alteraciones en el medio de cultivo pueden inducir a la liberación de la PspA de la superficie celular (Yother y White, 1994). Recientemente se ha demostrado que la delección parcial de los motivos contenidos en el dominio de unión a colina de la amidasa LytA, provoca una reducción secuencial de su actividad hidrolítica y su capacidad de unión (García y cols., 1994). Aunque sólo son necesarias cuatro repeticiones para mantener la interacción específica de la

LytA con la DEAE-celulosa, este número de motivos podría no ser suficiente para mantener la CspA unida fuertemente a la envuelta celular, permitiendo una liberación parcial de la proteína al medio de cultivo. Esto además explicaría el hecho de que la proteína CspC fuera más abundante en la fracción correspondiente a las envueltas celulares que en el sobrenadante (Fig. 10) ya que esta proteína tiene 5 motivos con lo cual la unión a la envuelta celular sería más intensa, liberándose menor cantidad al medio de cultivo. Si este comportamiento forma parte de un mecanismo regulador que responde a condiciones fisiológicas intracelulares o extracelulares es algo que debe ser elucidado.

La proteína asociada a la agregación de plaquetas (PAAP) de *S. sanguis* es una glicoproteína que contiene un 39% de ramnosa. Se ha sugerido que esta proteína secretable podría co-transportar, a través de la membrana, los polisacáridos ligados covalentemente que pueden posteriormente ser rotos y transferidos por las glucosiltransferasa a la pared celular (Erickson y Herzberg, 1993). La naturaleza glicosídica de la proteína CspA y la preponderancia de ramnosa podría sugerir una función en el transporte de polisacáridos. No ha sido posible demostrar, utilizando glicosidasas específicas, la naturaleza del enlace glicosídico. Sin embargo, Messner y cols. (1992) han descrito una proteína de superficie en *C. thermohydrosulfuricum*, con una unión *O*-glicosídica que no se afecta por el tratamiento con *O*-glicosidasa.

Podemos concluir, por tanto, que las proteínas Csp, independientemente de la función que puedan desempeñar, son proteínas asociadas a la pared a través de su dominio C-terminal, capaz de reconocer los residuos de colina presentes en los ácidos teicoicos de pared. El hecho de que constituyan las proteínas mayoritarias del medio de cultivo y que los genes que las codifican estén presentes en varias copias en el cromosoma bacteriano, nos hace pensar que deben desempeñar una función importante para la viabilidad del organismo.

2- CARACTERIZACIÓN MOLECULAR DE LA PROTEÍNA PcpA DE *S. pneumoniae*

El interés por clonar la segunda enzima lítica de neumococo, con actividad glucosaminidasa dependiente también de colina, nos condujo a la clonación de un gen codificante para una nueva proteína con motivos de unión a colina. Esta nueva proteína se denominó PcpA, y se demostró, mediante la obtención de un mutante defectivo en dicha proteína y por la expresión de dicho gen en el sistema homólogo, que no se trataba de la glucosaminidasa sino de una proteína con otras características.

La proteína PcpA presenta una estructura primaria en la que se pueden distinguir claramente dos dominios bien diferenciados. La región C-terminal comprende los últimos 239 aminoácidos y está constituida por 12 motivos repetidos de 20 aminoácidos de longitud e idénticos en su composición de aminoácidos (Fig. 35A). Estos motivos son homólogos a los motivos de unión a colina presentes en las enzimas líticas de neumococo y sus bacteriófagos (García y cols., 1988), en la PspA (Yother y Briles, 1992a) y en las proteínas Csp de *C. acetobutylicum* (este trabajo). El dominio N-terminal está constituido por motivos repetidos similares al de otras proteínas con motivos ricos en leucina.

Las repeticiones ricas en leucina (LRRs) se descubrieron por vez primera en la glicoproteína α_2 rica en leucina, también denominada LRG, una proteína del suero humano de función desconocida (Takahashi y cols., 1985). Las proteínas que contienen LRRs representan actualmente un grupo diverso de moléculas con diferente función y localización celular, presentes en una gran variedad de organismos y formando una superfamilia (para una revisión consultar Kobe y Deisenhofer, 1994). Las LRRs habitualmente se localizan en *tandem* variando el número de motivos entre 1, como es el caso de la glicoproteína Ib β de las plaquetas, a 30 como ocurre en la chaoptina. El dominio N-terminal de la proteína PcpA, que comprende los primeros 469 residuos, consta de 10 repeticiones de este tipo, de 22 a 26 aminoácidos de longitud, ricas en leucina y otros aminoácidos hidrofóbicos como la valina, fenilalanina e isoleucina. La longitud más común de estas repeticiones es de 24 residuos, aunque puede variar entre 20 y 29. Estas repeticiones se distribuyen en la PcpA en dos bloques de cinco motivos cada uno, comprendidos entre los residuos 143-260 y 293-409, respectivamente y unidos por lo que podría ser un péptido de interconexión o un motivo degenerado (Fig. 35C). Esta organización en bloques ya se ha descrito en otras

proteínas de esta superfamilia como es el caso de la proteína Toll de *Drosophila* (Hashimoto y cols., 1988), aunque en la mayor parte de los casos se agrupan constituyendo un único bloque (Kobe y Deisenhofer, 1994). A pesar de que las LRRs se distinguen fácilmente por su secuencia consenso siendo el aminoácido predominante la leucina, sin embargo, los métodos de comparación de secuencias han hecho pasar por alto muchas de estas proteínas debido a pequeñas deleciones, a la variación en longitud, a su composición de aminoácidos y a la presencia de secuencias consenso degeneradas. De hecho, cuando comparamos el dominio N-terminal de la proteína PcpA con la base de datos del EMBL, sólo apareció homología con las LRRs de la chaoptina y de los receptores de las hormonas gonadotrópicas, quizás por su mayor homología con los aminoácidos no consenso de estos motivos lo que nos permite hacernos una idea sobre la posible funcionalidad de esta proteína.

En general, aunque estas repeticiones están presentes en proteínas de función y origen diferente, sin embargo todas ellas parecen intervenir en procesos de interacción proteína-proteína, como es el caso de la chaoptina, y al menos la mitad de ellas, actúan como transductores de señales, como ocurre con los receptores de las hormonas gonadotrópicas. La característica más relevante de los motivos ricos en leucina es la alternancia de residuos hidrofóbicos e hidrofílicos siguiendo un patrón característico, lo que confiere a cada repetición un carácter anfipático. Ya se había postulado anteriormente que motivos anfipáticos mediaban las interacciones membrana-proteína o proteína-proteína, como ocurre con el segmento en hélice α de las cremalleras de leucina que intervienen en la dimerización de proteínas (Landschulz y cols., 1988) ó la hélice α anfipática formada por las repeticiones de la apolipoproteína A-I y A-II que permiten la interacción con lípidos (Segrest y col., 1974; Fukushima y cols., 1979; Sparrow y Gotto, 1982). En el caso de las LRRs se había propuesto también su intervención en las interacciones proteína-lípido (Takahashi y cols., 1985; Kataoka y cols., 1985; Reinke y cols., 1988) y proteína-proteína (Hofsteenge y cols., 1988; Kranz y Zipursky, 1990). La reciente cristalización del inhibidor de la ribonucleasa del hígado de cerdo (RI) ha permitido analizar la estructura tridimensional de las LRRs y estudiar el mecanismo por el cual estas repeticiones pueden ser responsables de la interacción con proteínas (Kobe y Deisenhofer, 1994). El RI es una proteína citoplásmica de 456 aminoácidos que contiene 15 LRRs de 29 (tipo A) y de 28 residuos (tipo B) cada una y que constituyen el 90% de la proteína. Estas LRRs están

constituídas por unidades estructurales β - α , con una pequeña cadena β y una hélice α . La hélice α se pliega de manera que queda antiparalela con la cadena β . A su vez las cadenas β procedentes de cada uno de los motivos repetidos se disponen formando una lámina β paralela. Debido a que el ancho de una hélice α es superior al espacio existente entre las cadenas β , estas proteínas adoptan una conformación no globular con forma de herradura quedando la lámina β en el arco interior de la herradura y las hélices α hacia el exterior (Fig. 55a). Los residuos consenso de estas repeticiones parecen desempeñar un papel estructural fundamental constituyendo el esqueleto principal de la proteína. El espacio existente entre los residuos de leucina es necesario para el correcto empaquetamiento de las cadenas β y las hélices α . El resto de los residuos consenso desempeñarían otras funciones. Estos autores piensan que este modelo puede ser aplicable a las LRRs de otras proteínas, lo que incluiría en nuestro caso, a la proteína PcpA, pues la mayoría de los residuos consenso de estos motivos que desempeñan un papel estructural, están conservados en las proteínas de esta superfamilia y, más en concreto, los residuos que se agrupan para constituir la lámina β (Fig. 55b). Como consecuencia, una de las propiedades estructurales de las LRRs responsable de la interacción con proteínas es su conformación no globular, haciendo que el área disponible para la interacción con proteínas globulares de menor tamaño se vea sustancialmente incrementada facilitando la interacción y mejorando la afinidad. Otra propiedad a tener en cuenta es la superficie expuesta al solvente de la lámina β paralela que podría favorecer también la interacción con las proteínas. Aunque estas características conformacionales sean responsables de la interacción proteína-proteína, la especificidad de la misma es probablemente debida a la composición específica de los residuos no consenso, aunque podría estar también influenciada por la longitud de las repeticiones y dominios flanqueantes.

La PcpA posee los aminoácidos consenso presentes en las LRRs de las proteínas descritas, sin embargo, sus aminoácidos no consenso parecen presentar sólo homología con las LRRs de la chaoptina y de los receptores de las hormonas gonadotrópicas. Esta homología nos permite sugerir que la proteína PcpA podría comportarse en *Neumococo* como una adhesina, permitiendo la interacción de este microorganismo con las células del huésped, favoreciendo de esta manera, la colonización de las vías altas del sistema

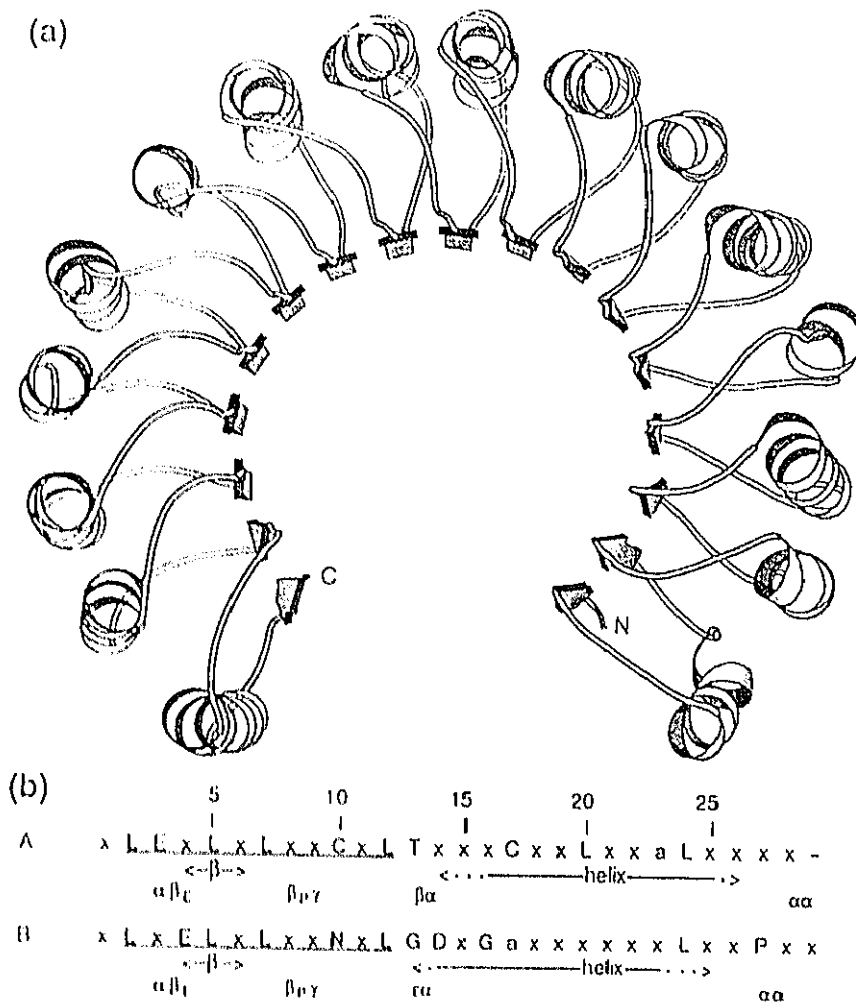


Figura 55- Estructura del inhibidor de la ribonucleasa del cerdo (RI). a) Diagrama de la estructura del RI. b) Secuencia consenso y estructura secundaria de las repeticiones ricas en leucina de tipo A y B presentes en dicha enzima. La región de las repeticiones, más conservada en toda la superfamilia de proteínas con LRRs, aparece subrayada. Debajo de dicha secuencia se indican los elementos de la estructura secundaria. La línea sólida marca la región principal de la lámina β o la hélice α ; los puntos señalan la extensión de la hélice α en diferentes repeticiones. Tomado de Kobe y Deisenhofer, 1994.

respiratorio. El hecho de que la chaoptina, una proteína de las células fotorreceptoras del ojo compuesto de la mosca y que está constituida por un 90% de repeticiones de este tipo, se clasifique como una molécula de adhesión celular (Krantz y Zipursky, 1990) y que la interacción de las hormonas gonadotrópicas con su receptor de membrana tenga lugar en las LRRs (Braun y cols., 1991) proporcionan soporte adicional a esta teoría.

En neumococo se ha descrito la existencia de dos mecanismos de adhesión a las células endoteliales. Un 60% de la adherencia parece estar asociada con alguna macromolécula no proteica de la superficie de la pared celular mientras que el 40% restante es sensible a tripsina, lo que hace pensar en un mecanismo mediado por proteínas (Cundell y Tuomanen, 1994). La fibronectina, una glicoproteína que juega un papel importante en la organización de las células en los tejidos eucariotas, contiene sitios de unión para distintas moléculas como proteoglicanos (Culp y cols., 1979), colágeno (Stenman y Vaheri, 1978) y para proteínas de varios microorganismos, en algunas de los cuales se han descrito los dominios de unión, como es el caso de *S. aureus* y *S. dysgalactiae* (McGavin, 1993) o de *S. pyogenes* (Talay y cols., 1994). También se ha descrito en neumococo la capacidad de unirse a la fibronectina (Andersson y cols., 1983), a la laminina y al colágeno tipo IV (Kostrzynska y Wadström, 1992). Aunque se ha indicado que en la adhesión de neumococo intervienen receptores disacáridicos de glicoconjugados presentes en las células del huésped, se sabe que la interacción con la fibronectina no tiene lugar a través de su fracción glucídica (Andersson y cols., 1983). Puesto que la mayor parte de los microorganismos que se unen a fibronectina son patógenos comunes, es posible que esta glicoproteína pudiera ser un sustrato clave en la adherencia bacteriana. El hecho de que existan proteoglicanos con LRRs capaces de unirse a diferentes componentes de la matriz extracelular como colágeno, fibronectina o laminina (Kobe y Deisenhofer, 1994), nos hace pensar que las LRRs de la proteína PcpA de neumococo desempeñarían una función similar, confiriendo a esta proteína un papel importante en los procesos de colonización. Por otro lado, la proteína PcpA no sería la primera adhesina descrita que presenta repeticiones de este tipo. Pocas son las proteínas bacterianas que contienen LRRs, concretamente cinco con la PcpA, de las cuales una de ellas posee propiedades adhesivas. Estas proteínas son además las únicas descritas en organismos procariotas. En *Listeria monocytogenes*, un microorganismo Gram-positivo clasificado como un patógeno intracelular facultativo (Mackaness, 1962), se han descrito dos de estas proteínas. Una de estas proteínas, la internalina o InlA se ha visto que promueve la adhesión de este microorganismo a las células del huésped. Es una proteína de superficie constituida por dos regiones con repeticiones denominadas A y B. La región A consta de 15 LRRs de 22 aminoácidos de longitud y la región B contiene tres repeticiones, dos de 70 aminoácidos y una de 49 aminoácidos diferente a la anterior. La región C-terminal posee los rasgos estructurales típicos de proteínas de superficie presentes en

bacterias Gram-positivas (Fischetti y cols., 1990). Este dominio comprende una región rica en glicina y prolina, el hexapéptido LPTTGD, una región transmembranal de 20 aminoácidos hidrofóbicos y una cola terminal cargada, de 7 aminoácidos. Se piensa que esta proteína se une a la membrana plasmática a través de su región C-terminal, mientras que la región N-terminal sería el dominio funcional en donde las repeticiones desempeñarían un papel primordial en la interacción de la internalina con las células del huésped (Gaillard y cols., 1991). Otra característica común que comparte el gen de la internalina con el gen de la proteína PcpA es la presencia de una secuencia palindrómica por encima de la caja promotora -35, la cual es similar a secuencias presentes en promotores que están bajo el control del activador pleiotrópico PrfA (Dramsí y cols., 1993). La proteína PrfA se sabe que regula positivamente la transcripción de un gran número de factores de virulencia en *L. monocytogenes* (Mengaud y cols., 1991; Wernars y cols., 1992). No sería de extrañar, por tanto, que un mecanismo similar regulara la transcripción del gen PcpA en neumococo, uniéndose a dicha secuencia palindrómica, una proteína similar a la PrfA que controlara tanto la expresión del gen *PcpA* como la de otros factores de virulencia de neumococo. Por detrás del gen *InlA*, se sitúa el gen *InlB* que contiene en su dominio N-terminal ocho LRRs de 22 aminoácidos cuya secuencia consenso es muy parecida a la de la *InlA* y cuya función se desconoce por el momento (Gaillard y col., 1991).

Recientemente se ha visto que la hemaglutinina filamentosa de *Bordetella pertusis*, bacteria Gram-negativa responsable de la tosferina, contiene repeticiones de 19 aminoácidos que, aunque no son estructuralmente iguales, sí presentan rasgos comunes con las LRRs descritas hasta ahora (Makhov y cols., 1994). Se trata también de una proteína de superficie que además de poseer un fuerte carácter inmunógeno, desempeña múltiples funciones de adhesión como la aglutinación de eritrocitos (Sato y cols., 1983) y la adhesión a varios tipos de células permitiendo la colonización del tracto respiratorio.

Respecto a las otras dos proteínas descritas con LRRs, destacan la YopM de *Y. pestis* (Leung y Straley, 1989) y la IpaH de *S. flexneri* (Hartman y cols., 1990). Ambas interlieren con la respuesta inflamatoria mediada por plaquetas, supuestamente debido a la homología existente entre las LRRs de estas proteínas y las de la glicoproteína Ib α (GPIb α) (Titani y cols., 1987). La GPIb α es uno de los dos receptores plaquetarios conocidos que unen el factor de von Willebrand, interacción esencial para conseguir una hemostasis normal al mediar la adhesión de plaquetas al subendotelio vascular y contribuir a la agregación

plaquetaria. Estas proteínas bacterianas parecen competir con la GPIb α por la unión del factor de von Willebrand permitiendo la adhesión de estos microorganismos al tejido dañado y facilitando su diseminación.

La naturaleza anfipática de las LRRs junto con la existencia de dos regiones hidrofílicas relevantes, no excluye la posibilidad de que el dominio N-terminal de la PcpA quede expuesto en la superficie celular, igual que ocurre con las proteínas de membrana que poseen LRRs (Kobe y Deisenhofer, 1994). En este caso, la proteína PcpA se comportaría como una proteína de superficie, con un mecanismo de unión a la membrana y/o anclaje a la pared celular de neumococo equivalente al de la proteína de superficie PspA (Yother y White, 1994), mediado por el dominio de unión a colina presente en su región C-terminal. Por otra parte, como ocurre con la internalina de *Listeria monocytogenes*, las LRRs de su dominio N-terminal serían responsables de la adhesión a las células del huésped a través de la interacción con la fibronectina u otras proteínas del tejido conjuntivo. La presencia del dominio de unión a colina permitiría la interacción de esta proteína con los residuos de colina presentes en el ácido lipoteicoico de la membrana plasmática o los ácidos teicoicos de pared quedando el dominio N-terminal expuesto en la superficie bacteriana. Esto vendría a reafirmar aún más la hipótesis de que la colina presente en la envoltura celular de neumococo parece servir de elemento de presión selectiva permitiendo, en el caso de proteínas de superficie como PcpA o PspA, diseñar un mecanismo de unión a la membrana y a la pared celular diferente al descrito para otras proteínas de superficie de microorganismos Gram-positivos (Fischetti y cols., 1990). La proteína PcpA, aunque posee una zona hidrofóbica al comienzo de la secuencia, no presenta la estructura típica de un péptido señal. De todas maneras, hay ejemplos descritos de proteínas secretables que carecen de péptido señal entre las que podríamos destacar la autolisina LytA, una proteína con dominio de unión a colina al igual que la PcpA (Díaz y cols., 1989).

La proteína PcpA es la que presenta mayor número de motivos repetidos entre las proteínas descritas con dominio de unión a colina. Cabe en consecuencia predecir que la interacción con los residuos de colina de la envoltura celular sería bastante fuerte, pues estudios con proteínas truncadas en varios de estos motivos, han demostrado que la interacción es dependiente del número de repeticiones en el extremo C-terminal (Yother y Briles, 1992a; García y cols., 1994). Tanto los motivos de las LRRs como los motivos de unión a colina son posiblemente el resultado de múltiples duplicaciones seguidas de

mutaciones acaecidas a lo largo de la evolución. En el caso de los motivos de unión a colina, el hecho de que sean idénticos en la composición de aminoácidos y prácticamente idénticos en la de nucleótidos, indica que estas duplicaciones han tenido lugar muy recientemente. Sin embargo, la variación en la longitud de las LRRs y la presencia de secuencias consenso específicas en las proteínas de esta superfamilia nos hace pensar que este dominio posee una historia evolutiva más compleja y rica que en el caso del dominio de unión a colina. Esto es fácil de entender si tenemos en cuenta que, en el caso del dominio de unión a colina, el único elemento de presión selectiva es la colina presente en la envoltura celular, mientras que en el caso de las LRRs va a depender del tipo de interacción en el que van a intervenir las LRRs.

La proteína PcpA es un ejemplo más de una proteína con estructura modular, como ocurre con todas las proteínas descritas con dominio de unión a colina, en donde este último es responsable de la unión de estas proteínas a los residuos de colina presentes en los ácidos teicoicos de pared ó en el ácido lipoteicoico de la membrana, siendo su función independiente del dominio N-terminal, posiblemente responsable en este caso de la adhesión. Esta estructura modular es también muy evidente en proteínas de la superfamilia de proteínas con LRRs, como es el caso de los receptores de las hormonas gonadotrópicas, las cuales presentan un dominio transmembranal propio de receptores acoplados a proteínas G y un dominio exoplásmico con motivos ricos en leucina responsable de la afinidad y selectividad de las hormonas gonadotrópicas (McFarlan y cols., 1989).

Debido a que la proteína PcpA puede jugar un papel importante en los procesos de adherencia y virulencia de neumococo, el conocimiento de su estructura y función permitiría una aplicación futura para la elaboración de vacunas, útiles en la prevención de infecciones producidas por este microorganismo.

3- CARACTERIZACIÓN DE UNA NUEVA SECUENCIA DE INSERCIÓN (IS1381) EN *S. pneumoniae*

En la presente Tesis se ha caracterizado una nueva secuencia de inserción, denominada IS1381, que precede a la secuencia correspondiente al gen *pcpA* pero que está codificada por la cadena complementaria. Las secuencias de inserción son elementos genéticos móviles al igual que los transposones. Un transposón es una secuencia de DNA capaz de abandonar su localización original y de insertarse en otro lugar del genoma. Los transposones están constituidos por una región central con genes que codifican para marcadores genéticos, como puede ser una resistencia a un antibiótico, flanqueada en cada extremo por una secuencia de inserción que permite su transposición. La secuencia de inserción (IS) lleva sólo los genes necesarios para su propia transposición junto con dos repeticiones invertidas en sus extremos de, al menos, 10 nucleótidos y dos repeticiones directas que son resultado de la duplicación de una corta secuencia del DNA en el lugar de la inserción (2 a 12 pb). Estos elementos contribuyen de manera significativa a la evolución de las especies por tres motivos: i) transponiendo secuencias de DNA que codifican para genes importantes (resistencias a antibióticos, metales etc.); ii) alterando el DNA del huésped mediante mutaciones por inserción, reorganizaciones en el cromosoma producidas por procesos de recombinación, y deleciones debidas a excisiones poco precisas; iii) transponiéndose a los plásmidos y dando lugar a transferencias de DNA mediadas por plásmidos.

En especies como neumococo donde los plásmidos son escasos y en donde algunos de los genes de resistencia a antibióticos residen en elementos cromosómicos móviles y otros surgen de la transferencia horizontal de DNA entre especies, la presencia de estos elementos despierta un gran interés aunque se desconoce por el momento el papel que desempeñan. En neumococo sólo se han descrito dos secuencias de inserción, la IS1202 (Morona y cols., 1994) y la IS1167 (Zhou y cols., 1995), a pesar de que ya se habían identificado con anterioridad elementos transponibles en forma de transposones conjugativos o simplemente como transposones. La IS1381 caracterizada en este trabajo presenta una estructura algo atípica, pues aunque tiene dos repeticiones directas de 7 nucleótidos, sin embargo, las posibles repeticiones invertidas son muy imperfectas (Fig. 32). Existen otros casos descritos de repeticiones imperfectas, como es el caso de la IS1167 de

S. pneumoniae (Zhou y cols., 1995), ó atípicas, poniendo como ejemplo la IS891 de la cianobacteria *Anabaena* sp. strain M-131 (Bancroft y Wolk, 1989). La IS1381 está constituida por dos ORFs, las cuales solaparían con que tuviera lugar un salto de fase de un nucleótido en el codón de terminación de la primera ORF (ORF3A; Fig. 32). Se ha observado que la expresión de un número importante de genes tanto en procariotas como en eucariotas se encuentra regulada a nivel traduccional mediante un salto de fase. La flexibilidad del código genético ofrece, por tanto, diferentes alternativas en la lectura del RNA mensajero (mRNA) con saltos programados en el marco de lectura en la dirección +1 ó -1 que no se deben confundir con los errores en la traducción. Este mecanismo de regulación se dice que está programado porque sucede en lugares específicos y depende de la secuencia nucleotídica del mRNA, estando implicadas ciertas secuencias, principalmente la del denominado "sitio de deslizamiento del ribosoma" que provoca el desplazamiento del RNA de transferencia (tRNA) sobre el mRNA. Hay que hacer notar que los casos mejor descritos que presentan un mecanismo de regulación de esta clase se han observado en las secuencias de inserción (Chandler y Fayet, 1993) y en genes de retrovirus, de algunas bacterias y de bacteriófagos (Atkins y cols., 1990). En el caso de las transposasas de las secuencias de inserción, aunque muchas son codificadas por un único marco de lectura, otras son codificadas por dos marcos de lectura situados en diferente fase pero solapantes tras la translocación del ribosoma durante la elongación y consecuente salto de fase (Chandler y Fayet, 1993). En nuestro caso podría suceder algo parecido puesto que la ORF3A y la ORF3B conjuntamente presentan una similitud muy significativa con la secuencia completa de la transposasa de la IS702 de la Cianobacteria *Calothrix* PCC7601 (Mazel y cols., 1991). Además, la IS1167 de *S. pneumoniae* descrita recientemente (Zhou y cols., 1995), presenta una organización similar a la IS1381, con dos ORFs solapantes en una base pero según un salto en el marco de lectura de tipo -1. Sin embargo, a diferencia de la IS1381, la ORF2 de la IS1167 equivalente a la ORF3A de la IS1381, presenta un RBS antes del codón de iniciación de la traducción lo que la hace susceptible de ser sintetizada de manera independiente a la ORF1.

En todos los casos estudiados con un salto de fase de tipo -1, se ha detectado la presencia de un heptanucleótido de tipo Y YYX XXXZ (Tabla 10) que, al parecer, es el que permite la translocación del mRNA, como se ha demostrado en algunos casos mediante

mutagénesis dirigida de los nucleótidos de dicha secuencia (Weiss y cols., 1989; Jacks y cols., 1988; Sekine y Ohtsubo, 1989; Escoubas y cols., 1991; Polard y cols., 1991).

Tabla 10- Elementos de la familia IS1 y miembros de la IS3

IS	Fuente	pb	Secuencia implicada en el salto de fase
IS1	<i>Escherichia coli</i>	768	T GAT-20-AAC ACG ATT TTC CGC <u>CAT TTA AAA AAC</u> TCA-4-TAA
IS2 (J01732)	<i>Escherichia coli</i>	1331	T GAA TAT GGA CGG <u>GCA AAA AAG</u> TGG-13-TAA
IS150 (X07037)	<i>Escherichia coli</i>	1443	T GAA-6-GAA ACG CGG CTT ATG TAT <u>CTA AAA AAG</u> CTG- 9-TGA
IS407 (M30378)	<i>Pseudomonas cepacia</i>	1236	T GAC-43-GAA GCG TTG GTC GTG GTC <u>AAG GGA AAG</u> CCC TGA
IS426 (X56562)	<i>Agrobacterium tumefaciens</i>	1319	T GAC-35-GCC CTT TCC AAA GCG GAC <u>TCA AAA AAA</u> CGG-13-TGA
IS600 (X05952)	<i>Shigella sonnei</i>	1264	T GAG CGA GAT ATA <u>TTA AAA AAA</u> GCA-13-TAA
IS861 (M22449)	<i>Streptococcus agalactiae</i>	1442	T GAG-2-AGA ACG GAG AAT GCC TTT <u>CTA AAA AAG</u> TTG-26-TAG
IS911 (X17613)	<i>Shigella dysenteriae</i>	1250	ATG GAG AAT GAA ATA <u>TTA AAA AAG</u> GCT-12-TAA
IS981 (M33933)	<i>Lactococcus lactis</i>	1222	T AAG GAA GAA AGA GAC ATC <u>TTA AAA AAA</u> GTA- 9-TGA
IS986 (X52471)	<i>Mycobacterium tuberculosis</i>	1354	T GAG-10- TTG CGA AGG GCG AAC <u>GCG ATT TTA AAG</u> ACC-13-TAA
ISR1 (X06616)	<i>Rhizobium lupini</i>	1259	T GAG-18- GCC CTG AAG GAC CTG CTG <u>GGA AAA AAG</u> TGG TGA
ISRm1 (X56563)	<i>Rhizobium meliloti</i>	1319	T AGC <u>CCA AAA AAA</u> CAT-14-TGA
IS3 (X02311)	<i>Escherichia coli</i>	1250	T GAA GAG CTG GCT ATC CTC <u>CAA AAG</u> GCC- 9-TGA
IS51 (M14365)	<i>Pseudomonas savastanoi</i>	1311	T AAG GCC TCA GCC <u>TAT TTT</u> GCC-10-TGA
IS476 (M28557)	<i>Xanthomonas campestris</i>	1226	GTT GGC TTC GGG <u>GTA AAA</u> CGC TAG
IS629 (X05953)	<i>Shigella sonnei</i>	1310	T AAC-6-TCC GCT <u>TAT TTT</u> GCG AAG GCG GAG <u>TTT GAC</u> CGC CTC TGG <u>AAA AAA</u> TGA
IS3411 (M19532)	<i>Escherichia coli</i>	1310	T AAC-6-TCC GCT <u>TAT TTT</u> GCG AAG GCG GAG TTC GAC CGC CTC TGG <u>AAA AAA</u> TGA

Los asteriscos hacen referencia a los elementos IS en los que se ha demostrado un salto de fase en el marco de lectura. El heptanucleótido, u otras secuencias implicadas en el salto de fase, se indican en negrita y subrayadas. Tomado de Chandler y Payet, 1993.

Se han encontrado también excepciones a la regla del heptanucleótido siendo a veces suficiente sólo cuatro nucleótidos iguales para promover un salto de tipo -1, especialmente si se localizan antes del codón de terminación (Tabla 10). Se ha visto que la frecuencia del salto de fase está modulada además por la presencia de estructuras secundarias a modo de

bucles o lazos detrás del heptanucleótido anterior y por la presencia de un RBS próximo al sitio de desplazamiento. En el caso de la transposasa de la *IS/381*, no cabe la posibilidad de un salto en el marco de lectura de tipo -1 pues, aunque existe un heptanucleótido a 11 pb del codón de terminación de la ORF3B que cumple los requisitos establecidos (Fig. 32), sin embargo, un salto de fase de este tipo supondría la aparición de un nuevo codón de terminación. Es más probable, por tanto, un salto de fase de tipo +1, el cual exige dos requerimientos: una pausa traduccional durante el proceso de elongación, debido a la presencia en el sitio aminoacil del ribosoma de un codon lento de descifrar ó de un codon de terminación, como ocurriría en nuestro caso, y la presencia de un RNA de transferencia especial ocupando el sitio peptidil del ribosoma. La naturaleza especial de este tRNA determinaría su habilidad para saltar a la posición +1 sobre el RNA mensajero en los casos en que sea posible ó bien provocaría que la incorporación del aminoacil-tRNA entrante quedese fuera de fase. Se ha observado que el tRNA del sitio aminoacil de la fase +1 así como el nucleótido que sigue al codón de la pausa son determinantes en el salto de fase (Pande y cols., 1995). Todo esto implica que el sitio de deslizamiento lo constituye un heptanucleótido al igual que en el salto de fase -1.

El sitio de deslizamiento de la *IS/381* estaría constituido por la secuencia GAA-UAA-U, diferente a los casos descritos hasta ahora (Pande y cols., 1995) (Fig. 32). Por otro lado el codón GAA no parece ser de los más fuertes para provocar un salto en el marco de lectura, como se demostró en el caso del factor de liberación 2 de *E. coli* (RF2 codificado por el gen *prfB*). Curran (1993) diseñó 32 variantes de una fusión *prfB/lacZ* mediante la sustitución del codon CUU, presente en el sitio de deslizamiento del gen *prfB*, por diferentes codones, valorando la eficiencia del salto de fase en función de la actividad β -galactosidasa resultante de la fusión. Una de las variantes llevaba el codon GAA presente en el sitio de deslizamiento de la *IS/381*, siendo el porcentaje del salto de fase de un 0,2% frente a un 48% para el codon CUU. Sin embargo, son pocas las proteínas descritas hasta ahora que están reguladas a nivel traduccional por un salto de fase +1, como son el retrotransposón Ty1 de *S. cerevisiae* (Mellor y cols., 1985) y el RF2 que utilizan el codón de deslizamiento CUU así como una anti-enzima de rata que posee el codón de deslizamiento UCC (Matsufuji y cols., 1995). Además, el uso de codones en *S. pneumoniae* no es igual que en *E. coli* por lo que no podemos hablar todavía de un mecanismo generalizado de regulación para el salto de fase +1. También parecen estar implicados en

cis, junto con el sitio de deslizamiento, los mismos elementos del mRNA que en el salto de fase -1 aunque, en los casos descritos, no se presentan conjuntamente, siendo en el caso del gen *prfB* una secuencia de Shine-Dalgarno (AGG-GGG) situada varios nucleótidos por encima del sitio de deslizamiento (Engelberg-Kulks y cols., 1994) y en el caso del gen de la anti-enzima de rata, la presencia de un bucle o lazo por detrás de dicho sitio. En el caso de la IS/381, sólo se observa una pequeña estructura de lazo con una energía libre -6.2 kcal/mol por detrás del codon de terminación.

Puesto que la transposasa de la IS702 de *Calothrix* PCC7601 está codificada por un único marco de lectura, la organización en dos marcos de lectura podría ser un rasgo específico de la IS/381 o podría reflejar simplemente una mutación en la secuencia de la copia analizada. Sin embargo, lo que sí parece claro es que la mutación no se produjo durante el proceso de clonación pues la secuencia del mismo inserto obtenida de otro clon diferente resultó ser exactamente la misma. El hecho de que existan dos ISs con características similares en el mismo microorganismo nos hace pensar más en la posibilidad de que su transposición, que en ambos casos es un hecho evidente puesto que se ha demostrado la existencia de varias copias en diferentes aislados de neumococo (Fig. 46), esté regulada por un mecanismo como el anteriormente descrito y no que sea debida a una mutación en la copia analizada. Esto tiene sentido si tenemos en cuenta tanto las propiedades de las ISs como el alto número de copias presentes en neumococo, que hace necesaria la existencia de un fuerte control y coordinación de sus actividades con el crecimiento bacteriano. La IS/381 sería, además, la primera secuencia de inserción regulada por un salto de fase en el marco de lectura de tipo +1 pues todas las ISs descritas hasta el momento presentan un salto de fase de tipo -1.

4- CARACTERIZACIÓN MOLECULAR DE LAS PROTEÍNAS PcpB y PcpC DE *S. pneumoniae*

La utilización de oligonucleótidos sintéticos degenerados, diseñados a partir de los aminoácidos más conservados en los motivos de unión a colina de las proteínas secuenciadas hasta el momento, nos ha permitido, por un lado, abordar el clonaje de los genes *pcpB* y *pcpC* codificantes para nuevas proteínas con motivos de unión a colina y, por otro, determinar si existen más genes que codifiquen para proteínas de este tipo.

Los genes *pcpB* y *pcpC* parecen constituir un operón al disponerse contiguos y al existir sólo 18 nucleótidos entre el codón de terminación del gen *pcpB* y el de iniciación del gen *pcpC*. La presencia de una posible secuencia promotora delante del codón de iniciación ATG del gen *pcpB* y, de un posible terminador detrás del gen *pcpC*, nos dan una idea del tamaño de la unidad transcripcional (Fig. 50). Ambos genes codifican para dos proteínas con dominios de unión a colina, según se deduce de su secuencia de aminoácidos.

La comparación con las proteínas de la bases de datos del EMBL sugiere una organización modular para la proteína PcpB, con un N-terminal que presenta los rasgos característicos de las proteínas de la familia de las serín-proteasas y un dominio C-terminal constituido por cuatro motivos repetidos, similares a los motivos de unión a colina descritos hasta ahora (Fig. 51A). La comparación de los 196 primeros residuos que constituyen el dominio N-terminal reveló una similitud de un 41% con la glutamil endopeptidasa de la cepa de *B. licheniformis* ATCC 14580 y un 50% de similitud con una serín-proteasa de *E. faecalis* (Figura 52). La glutamil endopeptidasa de *B. licheniformis* se sintetiza como una preproenzima de 316 aminoácidos que sufre un proceso de maduración con la pérdida de los primeros 94 residuos (Kakudo y cols., 1992). La similitud de esta proteasa con el dominio N-terminal de la proteína PcpB comienza en el aminoácido 25 de la proteína madura y continúa hasta el final de la misma. Esto parece indicar, que en la proteína PcpB, el dominio N-terminal correspondiente a una serín-proteasa, se ha fusionado con un dominio C-terminal de unión a colina creándose de esta manera una proteína quimérica potencialmente activa en neumococo.

La posibilidad de que la proteína PcpB sea una proteasa no se establece simplemente por su similitud con las proteasas antes mencionadas sino porque contiene, además, la estructura característica del centro activo de la familia de las serín-proteasas. En la figura 56

se muestra la organización del centro activo presente en seis serín-proteasas, incluyendo la PcpB. Se puede observar cómo los tres residuos claves de la quimiotripsina, His⁵⁷, Asp¹⁰² y Ser¹⁹⁵, se corresponden con la His²², Asp⁷⁵ y Ser¹⁴⁷ de la proteína PcpB, respectivamente.

	*	*	*
PcpB	LITAAHN	EYDLALL	GSSGSTV
BLase	VATAGHC	NYDYGAI	GQSGSPV
SprE	IVTNNHV	NADIAVV	GQSGSPI
Proteasa V8	VVDATHG	EGDLAIV	GNSGSPV
Tripsina	VVSAGHC	NNDIMLI	GDSGGPV
Quimiotripsina	VVTAHC	NNDITLI	GDSGGPL

Figura 56- Alineamiento de los aminoácidos constituyentes del sitio catalítico de varias serín-proteasas. BLase, glutamil endopeptidasa de *B. licheniformis* ATCC 14580; Proteasa V8 es la endopeptidasa de *S. aureus* V8; SprE hace referencia al producto del gen *sprE* de *E. faecalis*. Los aminoácidos marcados en negrita son aquellos que se encuentran conservados, como mínimo, en el 66,6% de los casos. Con un asterisco se señalan los aminoácidos clave del sitio catalítico.

La presencia de los residuos conservados antes mencionados nos permite clasificar provisionalmente a la proteína PcpB como una serín-proteasa, pudiendo incluirla, al igual que la glutamil endopeptidasa de *B. licheniformis*, dentro de la familia de la endopeptidasa α -lítica de *Lysobacter enzymogenes* (Silen y cols., 1988), perteneciente a la superfamilia de la quimiotripsina (Rawlings y Barret, 1994).

A principio de siglo, Rosenbow (1912) y Avery (1920) mencionaron la presencia de actividad proteolítica en extractos de neumococo. Posteriormente se ha descrito una proteasa IgA1 específica (Plaut, 1983) y otra que digiere IgG y IgM (Wikstrom y cols., 1984). Más recientemente, Courtney (1991) ha estudiado la actividad proteolítica de los sobrenadantes de cultivos de neumococo, incubados en un medio con colina al 1%, con el fin de evitar la lisis de este microorganismo. De esta manera, detectó la presencia de dos tipos de serín-proteasas, dependientes e independientes de Ca^{2+} , diferentes claramente a las anteriores. Preparaciones parcialmente purificadas de estas enzimas exhibían especificidad

por ciertos sustratos como la fibronectina, laminina, fibrinógeno, fibrina y elastina. La degradación de proteínas estructurales, de membrana o del tejido conjuntivo, puede favorecer la penetración de este microorganismo a través de las superficies de la mucosa y permitir el acceso del microorganismo al torrente sanguíneo. Además, la degradación de la fibronectina y otras proteínas tisulares puede debilitar la estructura del tejido alveolar provocando un incremento en su permeabilidad y desencadenar la formación de un edema, algo frecuente en la neumonía. Todos estos datos sugieren que estas enzimas contribuyen de manera importante en la patogénesis de las infecciones neumocócicas. La proteína PcpB podría muy bien tratarse de una de las serín-proteasas descritas por Courtney (1991), más si tenemos en cuenta que estas proteasas se obtuvieron de cultivos que contenían colina al 1%. La proteína PcpB debería estar unida a los residuos de colina de los ácidos teicoicos de pared o lipoteicoicos de la membrana, a través de su dominio C-terminal de unión a colina. La adición de colina al medio de cultivo favorecería la liberación de PcpB de la envuelta celular, determinando su aparición en el sobrenadante como ya se ha descrito en otras proteínas como la PspA (Yother y White, 1994). De esta manera, tras la adhesión de neumococo a células epiteliales o proteínas tisulares, como podría ocurrir a través de la proteína PcpA comentada en la discusión anterior, intervendrían enzimas proteolíticas, como la PcpB, permitiendo la colonización y diseminación de dicho microorganismo.

Se ha observado que el dominio C-terminal de la proteasa SGPC de *Streptomyces griseus* es homólogo al dominio de unión a quitina de las quitinasas A1 y D de *B. circulans*, por lo que se ha implicado a esta proteína en la degradación de proteínas unidas a la quitina, de manera que la adquisición de este dominio supondría una mejora en la especificidad de la enzima (Sachdev y cols., 1994). Algo similar podría ocurrir con la proteína PcpB tras la adquisición del dominio de unión a colina. Por otro lado, la presencia de repeticiones en el extremo C-terminal no sería algo exclusivo de la proteína PcpB pues las glutamil endopeptidasas V8 y Spase de *S. aureus* presentan 12 y 19 repeticiones respectivamente del tipo Pro-Asx-Asn ó Pro-Asx-Glx cuya función se desconoce por el momento (Yoshikawa y cols., 1992). Es posible que estas repeticiones desempeñen en *S. aureus* un papel semejante al de los motivos de unión a colina en neumococo.

La posibilidad de que la proteína PcpB actúe como una enzima lítica, hidrolizando las cadenas peptídicas que forman parte del peptidoglicano, es menos probable. Aunque desconocemos la especificidad sobre el tipo de enlace peptídico que va a hidrolizar no se

descarta que esta proteína se comporte como una glutamil endopeptidasa dada su similitud con la de *B. licheniformis*. De un modo u otro, si esta proteína interviene en los mecanismos de patogenicidad bacteriana o en el metabolismo de la pared celular, es algo que queda por resolver.

Respecto a la proteína PcpC, parece estar constituida toda ella por motivos repetidos que se extienden a lo largo de la proteína (Figura 51B) al igual que ocurría con las proteínas de unión a glucanos (GBPs). Sin embargo, sólo en el extremo C-terminal aparecen tres motivos similares y de igual longitud a los motivos de unión a colina junto con una cola terminal reminiscente de un cuarto motivo. El resto de la proteína se podría estructurar en motivos que han sufrido importantes deleciones a lo largo del proceso evolutivo pero conservando algunos de los aminoácidos consenso presentes en los motivos de unión a colina. Es una proteína predominantemente hidrofílica y presenta en su extremo N-terminal la secuencia característica de un péptido señal. El hecho de que esta proteína se transcriba conjuntamente con la PcpB nos hace pensar que quizás esta proteína esté implicada en su regulación puesto que es bien sabido que la actividad de las serín-proteasas está controlada por la presencia de inhibidores que impiden una acción incontrolada de dichas enzimas (Carrell y cols., 1987). El precursor de la proteasa IgA de *Neisseria gonorrhoeae* presenta un dominio C-terminal que resulta necesario para la secreción extracelular de dicha proteína (Pohlner y cols., 1987). Dada la ausencia de péptido señal en la proteína PcpB y su supuesta actividad endopeptidasa se podría pensar que quizás también la proteína PcpC favorezca la secreción de la misma, inhibiendo a la vez su actividad durante el proceso de translocación.

5- EL DOMINIO DE UNIÓN A COLINA

La clonación de los genes *csp* de *C. acetobutylicum* y, los genes *pcpA*, *pcpB* y *pcpC* de *S. pneumoniae*, nos ha permitido corroborar la existencia de un dominio de unión a colina conservado en proteínas con diferentes funciones y presentes en microorganismos filogenéticamente separados. La presencia de motivos repetidos se atribuye a la existencia de un motivo ancestral que ha sufrido duplicaciones y mutaciones diversas a lo largo del proceso evolutivo. Como se puede observar en la figura 57, donde se muestran los motivos de unión a colina de las diferentes proteínas identificadas hasta la fecha, el número de motivos es variable de unas proteínas a otras, siendo el mínimo para la proteína CspA y el máximo para la proteína PcpA. García y cols. (1994) habían demostrado que eran necesarios cuatro motivos en la enzima LytA para el reconocimiento eficiente de los residuos de colina de la pared celular, el mismo número que está presente en la proteína CspA. Por otro lado, la proteína CspA es más abundante en el sobrenadante del medio de cultivo que en las envueltas celulares, lo contrario a lo que ocurre con la proteína CspC (Fig. 10), lo que supone la existencia de más sitios de unión a la colina tras la adquisición de un quinto motivo y por ello un anclaje más intenso a la envoltura celular, impidiendo su liberación al medio de cultivo. La presencia de 12 motivos repetidos en la proteína PcpA supondría un anclaje mucho más fuerte a la envuelta celular. Esta proteína se caracteriza por ser la única con doce motivos idénticos en la composición de aminoácidos y casi idénticos en la de nucleótidos, lo que implica que las duplicaciones de dichos motivos han tenido lugar recientemente. Los motivos más degenerados son los presentes en las enzimas líticas de neumococo y sus bacteriófagos, lo que indica que han divergido mucho más que los presentes en las proteínas PspA, PcpA y PcpC, con motivos mucho más conservados.

Respecto a los aminoácidos más conservados, hay que destacar la conservación de la glicina en posición 16 en el 100% de los motivos, sugiriendo la importancia de este aminoácido en la estructura del dominio de unión a colina. Se encuentran también muy conservados el triplete de aminoácidos triptófano-tirosina-tirosina (posiciones 9,10 y 11) y el triptófano presente en la posición 2, los cuales en su defecto se encuentran sustituidos por otro aminoácido aromático. Cabe mencionar también la glicina en posición 1 y la metionina en posición 18, presentes en el 76% y 77% de los motivos, respectivamente. La presencia de aminoácidos tan conservados determina, posiblemente, la conformación de dicho

dominio, permitiendo la interacción con los residuos de colina presentes en las envueltas celulares.

	5					10					15					20				
CSPA	G	W	A	Q	T	S	A	G	W	T	Y	V	K	A	D	G	T	K	A	T
	G	W	L	Q	D	G	G	A	W	Y	Y	L	K	A	D	G	T	M	A	T
	G	W	I	Q	D	G	A	T	W	Y	Y	L	N	G	S	G	A	M	Q	T
	G	W	L	N	D	N	G	T	F	Y	Y	L	N	G	S	G	A	M	L	S
CSPB	G	W	V	Q	A	S	D	G	S	W	S	YI	K	A	D	G	N	K	A	T
	G	W	V	Q	D	G	S	T	W	Y	Y	F	K	A	D	G	S	M	A	T
	G	W	V	Q	D	G	S	T	W	Y	Y	F	Q	S	W	G	G	M	Q	T
	G	W	L	N	N	N	G	T	W	Y	Y	L	Q	S	W	G	G	M	K	T
	G	W	L	N	D	N	G	N	W	Y	Y	L	Q	S	W	G	G	M	Q	T
	G	W	F	N	D	N	G	T	W	Y	Y	L	Y	S	N	G	V	MA	A	NT
CSPC	G	W	V	Q	A	A	D	G	T	W	T	Y	N	KE	D	G	T	K	A	T
	G	W	L	N	L	N	G	V	W	Y	Y	L	N	T	D	G	V	M	A	T
	G	W	L	N	L	N	G	V	W	Y	Y	L	N	P	S	G	A	M	A	T
	S	W	L	N	L	N	G	T	W	Y	Y	L	N	S	S	G	A	M	A	T
	G	W	L	N	D	N	G	T	W	Y	Y	L	N	S	S	G	A	M	L	Y
CSPD	G	W	V	K	T	D	A	G	W	T	F	Y	N	T	D	P	T	Q	V	K
	GQ	W	V	N	D	G	G	V	W	Y	Y	I	K	A	D	G	T	M	A	T
	G	W	I	Q	D	G	S	T	W	Y	Y	L	Q	S	S	G	A	M	K	T
	G	W	L	N	D	N	G	T	W	Y	Y	L	Q	S	S	G	A	M	S	T
	G	W	L	N	D	N	G	T	W	Y	F	L	N	S	S	G	A	M	L	A
LYTA	G	W	Q	K	N	D	T	G	Y	W	Y	V	H	S	D	G	S	Y	PK	D
	K	F	E	K	I	N	G	T	W	Y	Y	F	D	S	S	G	Y	M	LA	D
	R	W	R	K	H	T	D	GN	W	Y	W	F	D	N	S	G	E	M	A	T
	G	W	K	K	I	A	D	K	W	Y	Y	F	N	E	E	G	A	M	K	T
	G	W	V	K	Y	K	D	T	W	Y	Y	L	D	A	KE	G	A	M	V	SN
	A	F	I	QS	AD	G	T	G	W	Y	Y	L	K	P	D	G	T	L	A	D
LYTA101	G	W	Q	Q	N	D	T	GT	W	Y	-	V	H	S	D	G	S	Y	P	KD
	K	F	E	K	V	N	G	T	W	Y	Y	F	D	G	S	G	Y	ML	A	D
	R	W	K	K	H	T	D	GN	W	Y	W	F	D	Q	S	G	E	M	A	T
	G	W	K	K	I	A	E	K	W	Y	Y	F	D	G	E	G	A	M	K	T
	G	W	V	K	Y	K	E	T	W	Y	Y	L	D	S	QN	G	D	M	V	SH
	A	F	V	K	S	N	D	G	W	Y	Y	L	K	E	D	G	T	I	A	
EJL	G	W	K	K	N	D	T	GT	W	Y	-	V	Y	S	D	G	S	Y	PK	D
	K	F	E	K	V	N	G	T	W	Y	Y	F	D	G	S	G	Y	M	LA	D
	R	W	K	K	H	S	D	GN	W	Y	W	F	D	Q	S	G	E	M	A	T
	G	W	K	K	I	A	E	K	W	Y	Y	F	D	V	E	G	A	M	K	T
	G	W	V	K	Y	K	D	T	W	Y	Y	L	D	S	KD	G	N	M	V	SN
	E	F	I	R	A	G	Q	G	W	Y	Y	I	K	P	D	G	T	M	A	
HBL	G	W	Q	K	N	G	T	G	Y	W	Y	V	H	S	D	G	S	Y	SK	D
	K	F	E	K	I	N	G	T	W	Y	Y	F	D	G	S	G	Y	M	L	SD
	R	W	K	K	H	T	D	GN	W	Y	Y	F	D	Q	S	G	E	M	A	T
	G	W	K	K	I	A	D	K	W	Y	Y	F	D	V	E	G	A	M	K	T
	G	W	V	K	Y	K	D	T	W	Y	Y	L	D	A	KE	G	A	M	V	SN
	A	F	I	O	S	A	DG	TG	W	Y	Y	L	K	P	D	G	T	L	A	

CPL1	T	W	K	Q	D	S	K	G	W	W	F	R	R	N	N	G	S	F	PY	N
	K	W	E	K	I	G	G	V	W	Y	Y	F	D	S	K	G	Y	C	L	TS
	E	W	L	K	D	N	E	K	W	Y	Y	L	K	D	N	G	A	M	A	T
	G	W	V	L	V	G	S	E	W	Y	Y	M	D	D	S	G	A	M	V	T
	G	W	V	K	Y	K	N	N	W	Y	Y	M	T	N	ER	G	N	M	V	SN
	E	F	I	K	S	G	K	G	W	Y	F	M	N	T	N	G	E	L	A	D
PSPA	G	W	K	Q	E	N	G	M	W	Y	F	Y	N	T	D	G	S	M	A	T
	G	W	L	Q	N	N	G	S	W	Y	Y	L	N	S	N	G	A	M	A	T
	G	W	L	Q	Y	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	A	K	V	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	L	Q	Y	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	A	K	V	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	L	Q	Y	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	A	K	V	N	G	S	W	Y	Y	L	N	A	N	G	A	M	A	T
	G	W	V	K	D	G	D	T	W	Y	Y	L	E	A	S	G	A	M	K	A
	SQ	W	F	K	V	S	D	K	W	Y	Y	V	N	G	L	G	A	L	A	
PCPA	G	W	V	K	D	K	G	L	W	Y	Y	L	N	E	S	G	S	M	A	T
PCPB	G	W	Q	N	L	G	N	K	W	Y	Y	L	R	S	S	G	A	M	A	T
	G	W	Y	Q	E	G	S	T	W	Y	Y	L	N	A	SN	G	D	M	K	T
	G	W	F	Q	V	N	G	N	W	Y	Y	A	Y	D	S	G	A	L	A	V
PCPC	G	W	K	K	I	N	G	S	W	Y	Y	Y	R	Q	H	-	D	K	Q	T
	G	W	Q	E	I	N	D	T	W	Y	Y	L	D	S	S	G	K	M	L	T
	D	W	Q	K	V	N	G	N	W	Y	Y	L	N	S	N	G	A	M	V	T
	G	S	Q	T	I	D	G	K	V	Y	N	F	A	S	S	G	E	W	I	
	G	W	V	K	D	N	G	T	W	Y	Y	L	N	S	S	G	A	M	A	T
			L	Q		K	D						D	D		S				
	76	87	32	56	34	35	52	30	92	91	84	57	43	27	44	100	35	77	57	66
			17	24		22	20						22		22		27			

Figura 57- Comparación de los dominios de unión a colina de diferentes proteínas.

CspA, CspB, CspC y CspD son las proteínas presentes en *C. acetobutylicum* NCIB 8052 descritas en este trabajo; LytA y LytA 101 son amidasas presentes en neumococo; Ejl, Hbl y Cpl1 son las enzimas líticas de los fagos EJ-1, HB-3 y Cp-1 que infectan neumococo; PspA es el antígeno de superficie de neumococo; PcpA, PcpB y PcpC son proteínas de neumococo caracterizadas en este trabajo, siendo la primera una posible adhesina y la segunda una serín-proteasa. Debajo se muestra la secuencia consenso de los motivos de unión a colina, considerando como aminoácidos más conservados los que se encuentran presentes en más del 15% de los casos. En negrita se muestran los aminoácidos conservados en más del 75% de los casos.

Por tanto, queda claro que la presencia de este dominio no es algo restringido a *S. pneumoniae*, como podría parecer en un principio, sino que está determinado por la presencia de colina en la envoltura celular de diferentes microorganismos. Esto explica la presencia de proteínas con dicho dominio en la estirpe de *C. acetobutylicum* NCIB 8052 y posiblemente en las estirpes NI-4 y NI-504 que presentan fragmentos de DNA que hibridan

con el fragmento que codifica para el dominio de unión a colina del gen *cspA*. Este aminoalcohol está actuando en estos microorganismos como un elemento de presión selectiva limitando la divergencia evolutiva natural del dominio de unión a dicho ligando. De esta manera, el dominio de unión a colina habría sido adquirido a lo largo de la evolución por proteínas funcionalmente diferentes, creándose proteínas quiméricas, con mayores ventajas al mejorar el sistema de reconocimiento del sustrato insoluble. El dominio N-terminal es responsable por tanto de la funcionalidad de estas proteínas, destacando las propiedades catalíticas en el caso de las enzimas líticas de neumococo y sus bacteriófagos, de las propiedades antigénicas como ocurre con la PspA, de las propiedades adhesivas como ocurriría con la PcpA y de las propiedades proteolíticas en el caso de la PcpB, mientras que el dominio C-terminal sería responsable del reconocimiento de la colina presente en la envoltura celular mejorando la especificidad de dichas proteínas e interviniendo posiblemente de manera indirecta en la regulación de las mismas. Las proteínas Csp de *C. acetobutylicum* presentan una organización similar aunque desconocemos por el momento su funcionalidad. La identificación de estas nuevas proteínas apoyan la teoría modular al estar constituidas por dos módulos que conservan su funcionalidad y ser a la vez intercambiables.

V- CONCLUSIONES

De los resultados presentados en esta memoria se derivan las siguientes conclusiones:

- 1- *C. acetobutylicum* NCIB 8052 posee al menos cuatro genes (*cspA*, *cspB*, *cspC* y *cspD*) que codifican para proteínas homólogas, cuyos dominios C-terminales presentan motivos de unión a colina.
- 2- Los genes *csp*, aunque no se localizan contiguos en el cromosoma bacteriano, sí se localizan relativamente próximos según los estudios de electroforesis en campo pulsado.
- 3- Los genes *cspA* y *cspC* son los únicos que se expresan en las condiciones ensayadas y, lo hacen como unidades transcripcionales independientes. Los altos niveles de expresión de dichos genes explicarían que las proteínas secretables CspA y CspC sean las más abundantes en el medio de cultivo.
- 4- Las proteínas CspA y CspC son glicoproteínas, habiendo sido detectada la presencia de un 12% de ramnosa (p/p) en la proteína CspA.
- 5- Las proteínas CspA y CspC son purificables en un solo paso por cromatografía de afinidad en columnas de DEAE-celulosa, característica que presentan las proteínas con dominios de unión a colina. La proteína CspA es la proteína nativa, descrita hasta la fecha, que presente un menor número de motivos repetidos en su dominio de unión a colina, lo que coincide con los datos obtenidos a partir de ensayos realizados con la amidasa LytA truncada en su dominio C-terminal.
- 6- Las proteínas Csp pueden considerarse como proteínas asociadas a la pared (WAPs), dada su presencia en las envueltas celulares, aunque se desconoce su función por el momento.
- 7- Las estirpes de *C. acetobutylicum* NI-4 y NI-504 contienen fragmentos de DNA que hibridan con el fragmento que codifica para el dominio de unión a colina de la proteína CspA de *C. acetobutylicum* NCIB 8052. Puesto que estas estirpes presentan colina en los

ácidos teicoicos de pared, dichos fragmentos podrían codificar para proteínas con dominios de unión a colina.

8- El diferente perfil de hibridación que presenta el DNA de las estirpes NCIB 8052 y NI-4, al utilizar como sonda el fragmento que codifica para el dominio de unión a colina de la proteína CspA, demuestra que no se trata de la misma cepa como habían considerado algunos autores.

9- El gen *pcpA* de *S. pneumoniae*, clonado a partir del fragmento de DNA que codifica para el dominio de unión a colina de la proteína CspA, codifica también para una proteína con dominio de unión a colina.

10- La presencia de repeticiones ricas en leucina (LRRs), implicadas en los procesos de interacción proteína-proteína, y presentes en el dominio N-terminal de la proteína PcpA, sugiere que esta proteína podría actuar como una adhesina. El hecho de que en el mecanismo de adherencia de neumococo esté implicado un componente de naturaleza proteica, proporciona soporte adicional a esta teoría.

11- Contiguo al gen *PcpA*, se ha caracterizado la secuencia de inserción *IS1381*. Su presencia, en múltiples copias, en diferentes cepas de laboratorio y aislados clínicos de *S. pneumoniae*, demuestra claramente que se trata de un elemento transponible.

12- Otros dos genes de *S. pneumoniae*, *pcpB*, y *pcpC*, clonados en este caso mediante la utilización de oligonucleótidos sintéticos degenerados diseñados en función de los aminoácidos más conservados en los motivos de unión a colina, codifican también para proteínas con dominios de unión a colina.

13- La similitud del dominio N-terminal de la proteína PcpB con algunas proteínas de la familia de las serín-proteasas, no descarta la posibilidad de una actividad proteolítica para esta proteína.

14- La conservación de la glicina en posición 16 en todos los motivos de unión a colina, sugiere que este aminoácido debe jugar un papel relevante para mantener la funcionalidad del dominio de unión a dicho ligando.

15- La organización modular de las proteínas con dominios de unión a colina, caracterizadas en este trabajo, aporta una mayor evidencia en favor de la teoría de la evolución modular y, expande el abanico de posibilidades funcionales.

16- La conservación del dominio de unión a colina en microorganismos que ocupan nichos ecológicos tan diferentes, como son *S. pneumoniae* y *C. acetobutylicum*, apoya la hipótesis de que la colina actúa como un fuerte elemento de presión selectiva, limitando la divergencia evolutiva de dicho dominio.

VI- BIBLIOGRAFÍA

- Andersson, B., Dahmén, J., Frejd, T., Leffler, H., Magnusson, G., Noori, G y Edén, C.S. 1983. Identification of an active disaccharide unit of a glycoconjugate receptor for pneumococci attaching human pharyngeal epithelial cells. *J. Exp. Med.* **158**: 559-570.
- Appleyard, R.K. 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics*, **39**: 440.
- Arondel, V., Benning, C. y Somerville, C.R. 1993. Isolation and functional expression in *Escherichia coli* of a gene encoding phosphatidylethanolamine methyltransferase (EC 2.1.1.17) from *Rhodobacter sphaeroides*. *J. Biol. Chem.* **21**: 16002-16008.
- Arrecubieta, C., López, R. y García, E. 1994. Molecular characterization of *cap3A*, a gene from the operon for the synthesis of the capsule of *Streptococcus pneumoniae* type3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. *J. Bacteriol.* **176**: 6375-6383.
- Atkins, J.F., Weiss, R.B. y Gesteland, R.F. 1990. Ribosome Gymnastics Degree of difficulty 9.5, style 10.0. *Cell* **62**: 413-423.
- Austrian, R. 1986. Some aspects of the pneumococcal carrier state. *J. Antimicrob. Chemoter.* **18**: 35-45.
- Avery, O.T. y Cullen, G.E. 1920. *J. Exp. Med.* **32**: 547-569.
- Avery, O.T., MacLeod, C. M. y McCarty, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. I. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79**: 137-158.
- Badger, E. 1944. The structural specificity of choline for the growth of type III pneumococcus. *J. Biol. Chem.* **153**: 183-191.
- Banas, J. A., Russell, R. R. B. y Ferretti, J. J. 1990. Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt. *Infect. Immun.* **58**: 667-673.
- Bancroft, I. y Wolk, C.P. 1989. Characterization of an insertion sequence (IS891) of novel structure from the cyanobacterium *Anabaena* sp. Strain M-131. *J. Bacteriol.* **171**: 5949-5954.
- Barany, F. y Tomasz, A. 1980. Genetic transformation of *Streptococcus pneumoniae* by heterologous plasmid deoxyribonucleic acid. *J. Bacteriol.* **144**: 698-709.
- Barany, F., Boeke, J.D. y Tomasz, A. 1982. Staphylococcal plasmids that replicate and express erythromycin resistance in both *Streptococcus pneumoniae* and *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **79**: 2991-2995.
- Barroso, L. A., Wang, S. Z., Phelps, C. J., Johnson, J. L. y Wilkins, T. D. 1990. Nucleotide sequence of *Clostridium difficile* toxin B gene. *Nucl. Acids, Res.* **18**: 4004.
- Bartlett, J.G., Taylor, N.S., Chang, T. y Dzink, J. 1980. Clinical and laboratory observations in *Clostridium difficile* colitis. *Am. J. Clin. Nutr.* **33**: 2521-2526.
- Behr, T., Fischer, Peter-Katalinic, W., y Egge, H. 1992. The structure of pneumococcal lipoteichoic acid. Improved preparation, chemical and mass spectrometric studies. *Eur. J. Biochem.* **207**: 1063-1075.
- Benner-Luger, D. y Boos, W. 1988. The *mglB* sequence of *Salmonella typhimurium* LT2; promoter analysis by gene fusions and evidence for a divergently oriented gene coding for the *mgl* repressor. *Mol. Gen. Genet.* **214**: 579-587.
- Berry, A. M., Lock, R. A., Hansman, D. y Paton, J. C. 1989. Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **57**: 2324-2330.
- Birnboim, H. C. y Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.
- Blake, C. C. F. 1978. Do genes in pieces imply protein in pieces?. *Nature.* **273**: 267.

- Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354**: 484-491.
- Bowles, L. K. y Ellefson, W. L. 1985. Effects of butanol on *Clostridium acetobutylicum*. *Appl. Env. Microbiol.* **50**: 1165-1170.
- Bracco, R.M., Krauss, M.R., Roe, A.S y MacLeod, M. 1957. Transformation reactions between pneumococcus and three strains of streptococci. *J. Exp. Med.* **106**: 247-259.
- Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.
- Braun, T., Schofield, P.R. y Sprengel, R. 1991. Amino-terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity. *EMBO J.* **10**: 1885-1890.
- Briese, T. y Hakenbeck, R. 1985. Interaction of the pneumococcal amidase with lipoteichoic acid and choline. *Eur. J. Biochem.* **146**: 417-427.
- Briles, E. B. y Tomasz, A. 1970. Radioautographic evidence for equatorial wall growth in a Gram-positive bacteria. *J. Cell. Biol.* **47**: 786-790.
- Brundish, D. E. y Baddiley, J. 1968. Pneumococcal C-substance, a ribitol teichoic acid containing choline phosphate. *Biochem. J.* **110**: 573-582.
- Carrell, R., y Boswell, D. R. 1987. In protease inhibitors, eds. Barret, A. y Salvesen, G. (Elsevier, Amsterdam), pp. 403-420.
- Courtney, H.S. 1991. Degradation of connective tissue proteins by serine proteases from *Streptococcus pneumoniae*. *Biochem. Biophys. Research. Comm.* **175**: 1023-1028.
- Crain, M. J., Waltman II, W. D., Turner, J. S., Yother, J., Talkington, D. K., McDaniel, L. S., Gray, B. M. y Briles, D. E. 1990. Pneumococcal surface protein A (PspA) is serologically variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infect. Immun.* **58**: 3293-3299.
- Croux, C. y García, J.L. 1991. Sequence of the *lyc* gene encoding the autolytic lysozyme of *Clostridium acetobutylicum* ATCC824: comparison with other lytic enzymes. *Gene.* **104**: 25-31.
- Croux, C. y García, J.L. 1992a. Reconstruction and expresión of the autolytic gene from *Clostridium acetobutylicum* ATCC 824 in *Escherichia coli*. *FEMS Microbiol Lett.* **95**: 13-20.
- Croux, C., Canard, B., Goma, G. y Soucaille, P. 1992b. Purification and characterization of an extracellular muramidase of *Clostridium acetobutylicum* ATCC 824 that acts on non-N-acetylated peptidoglycan. *Appl. Environ. Microbiol.* **58**: 1075-1081.
- Croux, C., Ronda, C., López, R. y García, J.L. 1993a. Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Mol. Microbiol.* **9**: 1019-1025.
- Croux, C., Ronda, C., López, R. y García, J.L. 1993b. Role of the C-terminal domain of the lysozyme of *Clostridium acetobutylicum* ATCC 824 in a chimeric pneumococcal-clostridial cell wall lytic enzyme. *FEBS.* **336**: 111-114.
- Culp, L., Murray, B.A. y Rollins, B.J. 1979. Fibronectin and proteoglycans as determinants of cell substratum adhesion. *J. Supramol. Struct.* **11**: 401-427.
- Cummins, C. S. y Johnson, J. L. 1971. Taxonomy of the Clostridia: wall composition and DNA homologies in *Clostridium butyricum* and other butyric acid-producing Clostridia. *J. Gen. Microbiol.* **67**: 33-46.
- Cundell, D.R. y Tuomanen, E.I. 1994. Receptor specificity of adherence of *Streptococcus pneumoniae* to human type-II pneumocytes and vascular endothelial cells *in vitro*. *Microbial. Pathogen.* **17**: 361-374.

- Curran, J.F. 1993. Analysis of effects of tRNA: message stability on frameshift frequency at the *Escherichia coli* RF2 programmed. *Nuc. Acid Res.* **21**: 1837-1843.
- Chandler, M. y Fayet, O. 1993. Translational frameshifting in the control of transposition in bacteria. *Mol. Microbiol.* **7**: 497-503.
- Datta, R. y Zeikus, J. G. 1985. Modulation of acetone-butanol-ethanol fermentation by carbon monoxide and organic acids. *Appl. Environ. Microbiol.* **49**: 522-529.
- Díaz, E. y García, J. L. 1990a. Construction of a broad-host-range pneumococcal promoter-probe plasmid. *Gene*, **90**: 163-167.
- Díaz, E., García, E., Ascaso, C., Méndez, E., López, R. y García, J.L. 1989. Subcellular localization of the major pneumococcal autolysin: A peculiar mechanism of secretion in *Escherichia coli*. *J. Biol. Chem.* **264**: 1238-1244.
- Díaz, E., López, R. y García, J. L. 1992. EJ-1, a temperate bacteriophage of *S. pneumoniae* with a *Myoviridae* morphotype. *J. Bacteriol.* **174**: 5516-5525.
- Díaz, E., López, R. y García, J.L. 1990b. Chimeric phage-bacterial enzymes: a clue to the molecular evolution of genes. *Proc. Natl. Acad. Sci. USA* **87**: 8125-8129.
- Díaz, E., López, R. y García, J.L. 1991. Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J. Biol. Chem.* **266**: 5464-5471.
- Doolittle, R. F. 1985. The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* **10**: 233-237.
- Dove, C.H., Wang, S.-Z., Price, S.B., Phelps, C.J., Lyerly, D.M., Wilkins, T.D. y Johnson, J.L. 1990. Molecular characterization of the *Clostridium* Toxin A gene. *Infect. Immun.* **58**: 480-488.
- Dower, W.J., Miller, J.F y Ragsdale, C.W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**: 6127-6145.
- Dramsí, S., Kocks, C. Forestier, C. y Cossart, P. 1993. Internalin-mediated invasion of epithelial cells by *Listeria monocytogenes* is regulated by the bacterial growth state, temperature and the pleiotropic activator *prfA*. *Mol. Microbiol.* **9**: 931-941.
- Dubos, R. J. 1937. *J. Exp. Med.* **66**: 101-112.
- Engelberg-Kulka, H. y Schoulaker-Schwarz, R. 1994. Regulatory implications of translational frameshifting in cellular gene expression. *Mol. Microbiol.* **11**: 3-8.
- Erickson, P.R. y Herzberg, M.C. 1993. Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. *J. Biol. Chem.* **266**: 5464-5471.
- Escoubas, J.M., Prere, M.F., Salvignol, I., Galas, D., Zerbib, D. y Chandler, M. 1991. Translational control of transposition activity of the bacterial insertion sequence IS1. *EMBO J.* **10**: 705-712.
- Fernández-Patrón, C., Castellanos-Serra, L y Rodríguez, P. 1992. Reverse staining of sodium dodecyl sulfate polyacrilamide gels by imidazol-zinc salts: sensitive detection of unmodified proteins. *BioTechniques* **12**: 564-573.
- Ferreti, J. J., Gilpin, M. L. y Russell, R. R. B. 1987. Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. *J. Bacteriol.* **169**: 4271-4278.
- Fischer, H. y Tomasz, A. 1984. Production and release of peptidoglycan and wall teichoic acid polymers in pneumococci treated with beta-lactam antibiotics. *J. Bacteriol.* **157**: 507-513.
- Fischer, H. y Tomasz, A. 1985. Peptidoglycan cross-linking and teichoic acid attachment in *Streptococcus pneumoniae*. *J. Bacteriol.* **163**: 46-54.

- Fischer, W., Behr, T., Hartmann, R., Peter-Katalinic, J. y Egge, H. 1993. Teichoic acid and lipoteichoic acid of *Streptococcus pneumoniae* possess identical chain structures. A reinvestigation of teichoic acid (C. polysaccharide). *Eur. J. Biochem.* **215**: 851-857.
- Fischetti, V.A., Pancholi, V. y Schneewind, O. 1990. Conservation of a hexapeptide sequence in the anchor region of surface protein from Gram-positive cocci. *Mol. Microbiol.* **4**: 1603-1605.
- Foster, S. J. 1993. Molecular analysis of the three major wall-associated proteins of *Bacillus subtilis* 168: evidence for processing of the product of a gene encoding a 285 kDa precursor two-domain ligand-binding protein. *Mol. Microbiol.* **8**:299-310.
- Friesen, J.D. y Tropak, M. 1983. Mutations in the *rplJ* leader of *Escherichia coli* that abolish feedback regulation. *Cell* **32**: 361-369.
- Fukushima, D., Kupferberg, J.P., Yokoyama, S. Kroon, D.J., Kaiser, E.T y Kézdy, F.J. 1979. *J. Am. Chem Soc.* **101**: 3703-3704.
- Furano, A. V. 1977. The elongation factor Tu coded by the *tufA* gene of *Escherichia coli* K-12 is almost identical to that coded by the *tufB* gene. *J. Biol. Chem.* **252**:2154-2157.
- Gaillard, J.L., Berche, P., Frehel, C., Gouin, E. y Cossart, P. 1991. Entry of *Listeria monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigen from Gram-positive cocci. *Cell.* **65**: 1127-1141.
- Galas, D.J. y Chandler, M. 1989. Bacterial insertion sequences. In "Mobile DNA" (D. E. Berg y M. M. Howe., Eds.), pp 109-162. American Society for Microbiology, Washington, DC.
- García, E., García, J. L., García, P., Arrarás, A., Sánchez-Puelles, J. M y López, R. 1988a. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. **85**: 914-918.
- García, E., García, J. L., Ronda, C., García, P. y López, R. 1985. Cloning and expression of the pneumococcal autolysin gene in *Escherichia coli*. *Mol. Gen. Genet.* **201**: 225-230.
- García, E., García, P. y López, R. 1993. Cloning and sequencing of a gene involved in the synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. *Mol Gen Genet* **239**:188-195.
- García, J. L., García, E., Arraras, A., García, P., Ronda, C. y López, R. 1987. Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. *J. Virol.* **61**: 2573-2580.
- García, J.L., Díaz, E., Romero, A. y García, P. 1994. Carboxy terminal deletion analysis of the major pneumococcal autolysin. *J. Bacteriol.* **176**: 4066-4072.
- García, J.L., García, E., Sánchez-Puelles, J. M. y López, R. 1988b. Identification of a lytic enzyme of *Clostridium acetobutylicum* that degrades choline-containing pneumococcal cell walls. *FEMS Microbiol. Lett.* **52**: 133-138.
- García, P., García, E., Ronda, C., López, R. y Tomasz, A. 1983a. A phage associated murein-hydrolase in *S. pneumoniae* infected with bacteriophage Dp-1. *J. Gen. Microbiol.* **129**: 489-497.
- García, P., García, E., Ronda, C., López, R., Jiang, R. Z. y Tomasz, A. 1986a. Mutants of *Streptococcus pneumoniae* that contain a temperature-sensitive autolysin. *J. Gen. Microbiol.* **132**: 1401-1405.
- García, P., García, E., Ronda, C., Tomasz, A. y López, R. 1983b. Inhibition of lysis by antibody against phage-associated lysin and requirement of choline-residues in the cell wall for progeny phage release in *Streptococcus pneumoniae*. *Curr. Microbiol.* **8**: 137-140.

- García, P., García, J. L., García, E., Sánchez-Puelles, J. M. y López, R. 1990. Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* 86: 81-88.
- García, P., García, J.L., García, E. y López, R. 1986b. Nucleotide sequence and expression of the pneumococcal autolysin gene from its own promoter in *Escherichia coli*. *Gene*. 43: 265-272.
- García, P., García, J.L., García, E. y López, R. 1989. Purification and characterization of the autolytic glycosidase of *Streptococcus pneumoniae*. *Biochem. Biophys. Res. Commun.* 158: 251-256.
- García, P., López, R., Ronda, C., García, E. y Tomasz, A. 1983c. Mechanism of phage-induced lysis in pneumococci. *J. Gen. Microbiol.* 129: 479-487.
- García, P., Méndez, E., García, E., Ronda, C. y López, R. 1984. Biochemical characterization of a murein hydrolase induced by bacteriophage Dp-1 in *Streptococcus pneumoniae*: comparative study between bacteriophage-associated lysin and the host amidase. *J. Bacteriol.* 159: 793-796.
- García-Bustos, J. F. y Tomasz, A. 1987. Teichoic acid-containing mucopeptides from *Streptococcus pneumoniae* as substrates for the pneumococcal autolysin. *J. Bacteriol.* 169: 447-453.
- García-Bustos, J. F. y Tomasz, A. 1989. Mechanism of pneumococcal cell wall degradation in vitro and in vivo. *J. Bacteriol.* 171: 114-119.
- García-Bustos, J. y Tomasz, A. 1990. A biological price of antibiotic resistance: major changes in peptidoglycan structure of penicillin-resistance pneumococci. *Proc. Natl. Acad. Sci. USA.* 87: 5415-5419.
- Gasc, A. M., Kauc, P., Barraillé, M., Sicard, M. y Goodgal, S. 1991. Gene localization, size and physical map of the chromosome of *Streptococcus pneumoniae*. *J. Bacteriol.* 173: 7361-7367.
- Gibson, T. J. 1984. Tesis Doctoral. Universidad de Cambridge, Inglaterra.
- Giffard, P. M., Allen, D. M., Milward, C. P., Simpson, C. L y Jacques, N. A. 1993. Sequence of the *gtfk* gene of *Streptococcus salivarius* ATCC 25975 and evolution of the *gtf* genes of oral streptococci. *J. Gen. Microbiol.* 139:1511-1522.
- Giffard, P.M., Simpson, C.L., Milward, C.P. y Jacques, N.A. 1991. Molecular characterization of a cluster of at least two glucosyltransferase genes in *Streptococcus salivarius* ATCC 25975. *J. Gen. Microbiol.* 137: 2577-2593.
- Gilbert, W. 1978. Why genes in pieces. *Nature.* 271: 501.
- Gilmore, K. S., Russel, R. R. y Ferretti, J. J. 1990. Analysis of the *Streptococcus downei gtfS* gene, which specifies a glucosyltransferase that synthesizes soluble glucans. *Infect. Immun.* 58: 2452-2458.
- Giudicelli, S. y Tomasz, A. 1984. Attachment of pneumococcal autolysin to wall teichoic acids, an essential step in enzymatic wall degradation. *J. Bacteriol.* 158: 1188-1190.
- Gottwald, M. y Gottschalk, G. 1985. The internal pH of *Clostridium acetobutylicum* and its effect on the shift from acid to solvent formation. *Arch. Microbiol.* 143: 42-46.
- Griffith, F. 1928. The significance of pneumococcal types. *J. Hyg.* 27: 113-159.
- Guarro, J., Cano, J., Leal, A., Gómez-Miranda, B. y Bernabé, M. 1993. Composition of the cell wall polysaccharides in some geophilic dermatophytes. *Mycopathologia* 122: 69-77.
- Hagen, P.O., Goldfine, H. y Williams, P.J. 1966. Phospholipids of bacteria with extensive intracytoplasmic membranes. *Science* 151: 1543-1544.

- Hanada, N., Isobe, Y., Aizawa, Y., Katayama, T., Sato, S. y Inoue, M. 1993. Nucleotide sequence analysis of the *gltT* gene from *Streptococcus sobrinus* OMZ176. *Infect. Immun.* 61: 2096-2103.
- Hanahan, D. 1983. Studies on transformation of *E. coli* with plasmids. *J. Mol. Biol.* 166: 557-580.
- Hartman, B., Venkatesan, M., Oaks, E.V. y Buysse, J.M. 1990. Sequence and molecular characterization of a multicopy invasion plasmid antigen gene, *ipaH*, of *Shigella flexneri*. *J. Bacteriol.* 172: 1905-1915.
- Hartmanis, M. G. N., Klason, T. y Gatenbeck, S. 1984. Uptake and activation of acetate and butyrate in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* 20: 66-71.
- Hashimoto, C., Hudson, K.L. y Anderson, K.V. 1988. The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell* 52: 269-279.
- Hofsteenge, J., Kieffer, B., Matthies, R., Hemmings, B.A. y Stone, S.R. 1988. Amino acid sequence of the ribonuclease inhibitor from porcine liver reveals the presence of leucine rich repeats. *Biochem.* 27: 8537-8544.
- Höltje, J. V. y Tomasz, A. 1974. Teichoic acid phosphorylcholine esterase. *J. Biol. Chem.* 249: 7032-7034.
- Höltje, J.V. y Tomasz, A. 1975a. Lipoteichoic acid: a specific inhibitor of autolysin activity in pneumococcus. *Proc. Nat. Acad. Sci. USA* 72: 1690-1694.
- Höltje, J.V. y Tomasz, A. 1975b. Specific recognition of choline residues in the cell wall teichoic acid by the N-Acetylmuramyl-L-alanine amidase of pneumococcus. *J. Biol. Chem.* 250: 6072-6076.
- Höltje, J.V. y Tomasz, A. 1976. Purification of the pneumococcal N-acetylmuramyl-L-alanine amidase to biochemical homogeneity. *J. Biol. Chem.* 251:4199-4207.
- Hollinghead, S. K., Fischetti, V. A., y Scott, J. R. 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol. Gen. Genet.* 207:196-203.
- Hollingshead, S. K., Readdy, T. L., Yung, D. L. y Bessen, D. E. 1993. Structural heterogeneity of the *emm* gene cluster in group A streptococci. *Mol. Microbiol.* 8:707-717.
- Honda, O., Kato, C. y Kuramitsu, H. K. 1990. Nucleotide sequence of the *Streptococcus mutans* *gltD* gene encoding the glucosyltransferase-S enzyme. *J Gen. Microbiol.* 136: 2099-2105.
- Hongo, M. 1960. US patent no. 2945786.
- Hongo, M. y Murata, A. 1965. Bacteriophages of *Clostridium sacharoperbutylaceticum*. *Agr. Biol. Chem.* 12: 1135-1139.
- Hongo, M., Tahara, Y. y Ogata, S. 1974. Purification and properties of bacteriophage HM-7 induced lytic enzyme of *Clostridium sacharoperbutylaceticum*. *Agr. Biol. Chem.* 38: 755-761.
- Horne, D., Hakenbeck, R. y Tomasz, A. 1977. Secretion of lipid by inhibition of peptidoglycan synthesis in Streptococci. *J. Bacteriol.* 132: 704-717.
- Howard, L. V. y Gooder, H. 1974. Specificity of the autolysin of *Streptococcus (Diplococcus) pneumoniae*. *J. Bacteriol.* 117: 796-804.
- Howard, L. V. y Gooder, H. 1974. Specificity of the Autolysin of *Streptococcus (Diplococcus) pneumoniae*. *J. Bacteriol.* 117: 796-804.
- Hutkins, R.W. y Kashket, E.R. 1986. Phosphotransferase activity in *C. acetobutylicum* from acidogenic and solventogenic phase of growth. *Appl. Environ. Microbiol.* 51: 1121

- Ingram, L. O. 1986. Microbial tolerance to alcohols: role of the cell membrane. *TIBETCH* 40.
- Inouye, S. e Inouye, M. 1985. Up-promoter mutations in the *lpp* gene of *Escherichia coli*. *Nucleic Acids. Res.* 13: 3101-3110.
- Jacks, T., Madhani, H.D, Masiarz, F.R y Varmus, H. 1988. Signals for ribosomal frameshifting in the rous sarcome virus Gag-Pol region. *Cell* 55: 447-458.
- Jones, D. T. y Woods, D. 1986. Acetone-butanol fermentation revisited. *Microbiol. Rev.* 50: 484-524.
- Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H., y Yoshida, N. 1992. Purification, characterization, cloning and expression of a glutamic acid-specific protease from *Bacillus licheniformis* ATCC 14580. *J. Biol. Chem.* 33: 23782-23788.
- Kataoka, T., Broek, D. y Wigler, M. 1985. DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate cyclase. *Cell*, 43: 493-505.
- Kawamura, T. y Shockman, G. D. 1983. Evidence for the presence of a second peptidoglycan hydrolase in *Streptococcus faecium*. *FEMS Microbiol. Lett.* 19: 65-69.
- Kilpper-Bälz, R., Wenzig, P. y Schleifer, K.H. 1985. Molecular relationships and classification of some viridans streptococci as *Streptococcus oralis* and emended description of *Streptococcus oralis* (Bridge and Sneath 1982). *Int. J. Syst. Bacteriol.* 35: 482-488.
- Kobe, B. y Deisenhofer, J. 1994. The leucine-rich repeat: a versatile binding motif. *TIBS* 19: 415-421.
- Kostrzynska, M., y Wadström, T. 1992. Binding of laminin, type IV collagen, and vitronectin by *Streptococcus pneumoniae*. *Zbl. Bakt.* 277: 80-83.
- Krantz, D.E. y Zipursky, S.L. 1990. *Drosophila* chaoptin, a member of the leucine rich repeat family, is a photoreceptor cell-specific adhesion molecule. *EMBO J.* 9: 1969-1977.
- Krivan, H.C., Clark, G.F., Smith, D.F. y Wilkins, T.D. 1987. Cell surface binding site for *Clostridium difficile* enterotoxin: evidence for a glyconjugate containing the sequence Gal α 1-Gal β 1-4GlcNAc. *Infect Immun* 53: 573-581.
- Kushner, S.R. 1978. En Genetic engineering (Boyer, H.B. y Nicosia, S., eds.) pp 17. Elsevier/North Holland, Amsterdam.
- Kyte, J. y Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105-132.
- Lacks, S. 1970. Mutants of *Diplococcus pneumoniae* that lack deoxiribonucleases and other activities possibly pertinent to genetic transformation. *J. Bacteriol.* 101: 373-83.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* 227: 680-685.
- Landschulz, W.H., Johnson, P.F. y McKnight, S. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240: 1759-1764.
- Leung, K.Y. y Straley, S.C. 1989. The *yopM* gene of *Yersinia pestis* encodes a released protein having homology with the human platelet surface protein GPIb α . *J. Bacteriol.* 171: 4623-4632.
- Lingren, P.-E., McGavin, M. J., Signäs, C., Guss, B., Gurusiddappa, S., Höök, M. and Lindberg, M. 1993. Two different genes coding for fibronectin-binding proteins from *Streptococcus dysagalactiae*. The complete nucleotide sequences and characterization of the binding domains. *Eur. J. Biochem.* 214: 819-827.
- López, C, Grubb, A. y Méndez, E. 1982a. Human protein HC display in its carboxyl-terminal amino acid sequence. *FEBS Lett.* 144: 349-353.

- López, J.A., Chung, D.W., Fujikawa, K., Hagen, F.S., Papayannopoulou, T. y Roth, G.J. 1987. Cloning of the α chain of human platelet glycoprotein Ib: a transmembrane protein with homology to leucine α_2 -glycoprotein. *Proc. Natl. Acad. Sci.* **84**: 5615-5619.
- López, R., García, J.L., García, E., Ronda, C. y García P. 1992. Structural analysis and biological significance of the cell wall lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *FEMS Microbiol. Lett.* **100**: 439-448.
- López, R., García, P., Ronda, C. y Tomasz, A. 1982b. Choline containing bacteriophages receptors in *Streptococcus pneumoniae*. *J. Bacteriol.* **151**: 1581-1590.
- López, R., Ronda, C., García, P., Escarmis, C. y García, E. 1984. Restriction cleavage maps of the DNAs of *Streptococcus pneumoniae* bacteriophages containing protein covalently bound to their 5' ends. *Mol. Gen. Genet.* **197**: 67-74.
- López, R., Sánchez-Puelles, J. M., García, E. García, J. L., Ronda, C. y García, P. 1986. Isolation, characterization and physiological properties of an autolytic-deficient mutant of *Streptococcus pneumoniae*. *Mol. Gen. Genet.* **204**: 237-242.
- Ludwig, W., Seewaldt, E., Kilpper-Balz, R., Schleifer, K. H., Magrum, L., Woese, C. R., Fox, G. E. y Stackebrandt, E. 1985. The phylogenetic position of *Streptococcus* and *Enterococcus*. *J. Gen. Microbiol.* **131**: 543-551.
- Lund, E. y Henriksen, J. 1978. Laboratory diagnosis, serology and epidemiology of *Streptococcus pneumoniae*. In: T. Bergan and J.R. Norris (ed.). *Methods in Microbiol.*
- Lyerly, D.M. y cols., 1990. Vaccination against lethal *Clostridium difficile* enterocolitis with a nontoxic recombinant peptide of toxin A. *Curr. Microbiol.* **21**: 29-33.
- Mackaness, G.B. 1962. Cellular resistance to infection. *J. Exp. Med.* **116**: 381-406.
- Makhov, A.M., Hannah, J.H., Brennan, M.J., Trus, B.L., Kocsis, E., Conway, J.F., Wingfield, P.T., Simon, M.N. y Steven, A.C. 1994. Filamentous hemagglutinin of *Bordetella pertussis*. A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. 1994. *J. Mol. Biol.* **241**: 110-124.
- Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F., Gesteland, R.F. y Hayashi, S. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell.* **80**: 51-60.
- Matta-El-Amouri, G., Janati-Idrissi, R., Assobhei, O., Petitdemange, H. y Gay, R. 1985. Mechanism of the acetone formation by *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* **30**: 11-16.
- Mazel, D., Bernard, C., Schwarz, R., Castets, A.M., Houmard, J. y Tandeau de Marsac, N. 1991. Characterization of two insertion sequences, IS701 and IS702, from the cyanobacterium *Calothrix* species PCC 7601. *Mol. Microbiol.* **5**: 2165-2170.
- McFarland, K.C., Sprengel, R.S., Phillips, H.S., Köhler, M., Rosemblit, N. Nikolics, K., Segaloff, D.L. y Seeburg, P.H. 1989. Lutropin-choriogonadotropin receptor: an unusual member of the G protein-coupled receptor family. *Science* **245**: 494-499.
- McGavin, M.J., Gurusiddappa, S., Lindgren, P-E., Lindberg, M., Raucci, G., y Höök, M. 1993. Fibronectin receptors from *Streptococcus dysagalactiae* and *Staphylococcus aureus*. *J. Biol. Chem.* **268**: 23946-23953.
- Mellor, J., Fulton, S.M., Dobson, W.W., Kingsman, S.M. y Kingsman, A.J. 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. *Nature* **313**: 243-246.
- Mengaud, J. Dramsi, S., Gouin, E., Vázquez-Boland, J.A., Milon, G., y Cossart, P. 1991. Pleiotropic control of *Listeria monocytogenes* virulence factors by a gene which is autoregulated. *Mol. Microbiol.* **5**: 2273-2283.

- Messner, P., Christian, R., Kolbe, J., Schulz, G. y Sleytr, U.B. 1992. Analysis of a novel linkage unit of O-linked carbohydrates from the crystalline surface layer glycoprotein of *Clostridium termohydrosulfuricum* S102-70. *J. Bacteriol.* 174: 2236-2240.
- Mollet, B., Knol, J., Poolman, B., Marciset, O. y Delley, M. 1993. Directed genomic integration, gene replacement, and integrative gene expression in *Streptococcus thermophilus*. *J. Bacteriol.* 175: 4315-4324.
- Moreira, A. R., Ulmer, D. C. y Linden, J. C. 1981. Butanol toxicity in the butylic fermentation. *Biotechnol. Bioeng. Symp.* 11: 567-579.
- Morgenroth, J. y Kaufmann, M. 1912. Arzneifestigkeit bei bakterien (pneumokokken). *Z. Immuniatatsforsch. Exp. Ther.* 15: 610-624.
- Morona, J.K., Guidolin, A., Morona, R., Hansman, D. y Paton, J.C. 1994. Isolation, characterization, and nucleotide sequence of IS1202, an insertion sequence of *Streptococcus pneumoniae*. *J. Bacteriol.* 176: 4437-4443.
- Mosser, J. L. y Tomasz, A. 1970. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *J. Biol. Chem.* 245: 287-298.
- Neufeld, F. 1900. Ueber eine spezifische bakterolytische wirkung der galle. *Z. Hyg. Infekt.* 34: 454-464.
- Neufeld, F. 1902. Ueber die agglutination der pneumokokken und über die theorieen der agglutination. *Z. Hyg. Infekt.* 40: 54-72.
- Ogata, S. y Motoyoshi, H. 1973. Bacterial lysis of *Clostridium* species. Lysis of *Clostridium* species by univalent cation. *J. Gen Appl. Microbiol.* 19: 251-261.
- Ogata, S., Tahara, Y. y Hongo, M. 1974. Mode of action of *Clostridium saccharoperbutylacetonicum* cell wall peptidoglycan. *Agr. Biol. Chem.* 38: 763-768.
- Ogata, S., Tahara, Y. y Hongo, M. 1975. Chemical composition of cell wall peptidoglycan from *Clostridium saccharoperbutylacetonicum* studied with phage endolysins and gas chromatography. *J. Gen. Appl. Microbiol.* 21: 65-74.
- Oultram, J.D., Loughlin, M., Swinfield, T.-J., Brehm, J.K., Thompson, D.E. y Minton, N.P. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiol. Lett.* 56: 83-88.
- Pande, S., Vimaladhithan, A., Zhao, H. y Farabaugh, P.J. 1995. Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. *Mol. Cell. Biol.* 15: 298-304.
- Paton, J. C., Boulnois, A. G. J. y Mitchell, T. J. 1993. Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: the role of pneumococcal proteins. *Annu. Rev. Microbiol.* 47: 89-115.
- Paul, G., Lottspeich, F. y Wieland, P. 1986. Asparaginyln-*N*-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan. *J. Biol. Chem.* 261: 1020-1024.
- Plaut, A.G. 1983. The IgA1 proteases of pathogenic bacteria. *Ann. Rev. Microbiol.* 37: 603-622.
- Podvin, L., Reyset, G., Hubert, J. y Sebald, M. 1988. Presence of choline in teichoic acid of *Clostridium acetobutylicum* NI-4 and choline inhibition of autolytic functions. *J. Gen. Microbiol.* 134: 1603-1609.
- Pohlner, J., Halter, R., Beyreuther, K. y Meyer, T.F. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature.* 325: 458-462.
- Polard, P., Prere, M.F., Chandler, M.F. y Fayet, O. 1991. Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. *J. Mol. Biol.* 222: 465-477.

- Price, S.B., Phelps, C.J., Wilkins, T.D. y Johnson, J.L. 1987. Cloning of the carbohydrate-binding portion of the toxin A gene of *Clostridium difficile*. *Curr. Microbiol.* 16: 55-60.
- Rane, L. y Subbarow, Y. 1940. Nutritional requirements of the pneumococcus 1. Growth factors for types I, II, V, VII, VIII. *J. Bacteriol.* 40: 695-704.
- Rawlings, N.D. y Barrett, A.J. 1994. Families of serine peptidases. *Methods in enzymology*, 244: 19.
- Reinke, R., Krantz, D.E., Yen, D. y Zipursky, L. 1988. Chaoptin, a cell surface glycoprotein required for *Drosophila* photoreceptor cell morphogenesis, contains a repeat motif found in Yeast and Human. *Cell.* 52: 291-301.
- Rogers, H.J. 1974. Peptidoglycan (muropeptides) structure, function and variations. *Ann. N. Y. Acad. Sci.* 235: 29-51.
- Rogers, P. 1986. Genetics and biochemistry of *Clostridium* relevant to development of fermentation processes. *Adv. Appl. Microbiol.* 31: 1-60.
- Romero, A., López, R. y García, P. 1990b. Sequence of the *Streptococcus pneumoniae* bacteriophage reveals high homology with the major host autolysin. *J. Bacteriol.* 172: 5064-5070.
- Romero, A., López, R. y García, P. 1990a. Characterization of the pneumococcal bacteriophage HB-3 amidase: cloning and expression in *Escherichia coli*. *J. Virol.* 64: 137-142.
- Ronda, C., García, J.L. y López, R. 1988. Characterization of genetic transformation in *Streptococcus oralis* NCTC 11427: expression of the pneumococcal amidase in *S. oralis* using a new shuttle vector. *Mol Gen Genet* 215:53-57.
- Ronda, C., García, J.L. y López, R. 1991. Teichoic acid choline esterase, a novel hydrolytic activity in *Streptococcus oralis*. *FEMS Microbiol. Lett.* 80: 289-294.
- Ronda, C., García, J.L., García, E., Sánchez-Puelles, J.M. y López, R. 1987. Biological role of the pneumococcal amidase. *Eur. J. Biochem.* 164: 621-624.
- Ronda, C., López, R. y García, E. 1981. Isolation and characterization of a new bacteriophage, Cp-1, infecting *Streptococcus pneumoniae*. *J. Virol.* 40: 551-559.
- Roos, J.W., McLaughlin, J.K. y Papoutsakis, E.T. 1985. The effect of pH on nitrogen supply, cell lysis and solvent production in fermentations of *Clostridium acetobutylicum*. *Biotechnol. Bioeng.* 27: 681-694.
- Rosenberg, M. y Court, D. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Ann. Rev. Gen.* 13: 319-53.
- Rosenbow, E.C. 1912. *J. Infect. Dis.* 11: 286-293.
- Sørensen U. B. S., Blom, J., Birch-Andersen, A. y Henrichsen, J. 1988. Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infection and Immunity.* 56: 1890-1896.
- Sørensen, U.B. y Blom, J. 1992. Capsular polysaccharide is linked to the outer surface of type 6A pneumococcal cell walls. *APMIS* 100: 891-893.
- Sørensen, U.B., Henrichsen, J. y Chen, H.C. 1990. Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. *Microb. Pathog.* 8: 325-34.
- Sachdev, S.S., Kalmar, G.B., Willis, L.G. y Borgford, J. 1994. *Streptomyces griseus* protease C. A novel enzyme of the chymotrypsin family. *J. Biol. Chem.* 269: 20167-20171.
- Saha, B. C., Lamed, R. y Zeikus, J. G. 1989. Clostridia. Clostridial enzymes. (Minton, N. G. y Clarke, D. J, eds.) pp-227-256. Plenum Press. New York.

- Sambrook, J., Fritsch, E.F. y Maniatis, T. 1989. Molecular cloning: A Laboratory manual. 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
- Sánchez-Puelles, J. M., Ronda, C., García, E., Méndez, E., García, J. L. y López, R. 1986a. A new peptidoglycan hydrolase in *Streptococcus pneumoniae*. FEMS Microbiol. Lett. **35**: 163-166.
- Sánchez-Puelles, J. M., Sanz, J. M., García, J. L. y García, E. 1990. Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. Gene. **89**: 69-75.
- Sánchez-Puelles, J.M., García, J.L., López, R. y García, E. 1987. 3'-End modifications of the *Streptococcus pneumoniae* *lytA* gene: role of the carboxi terminus of the pneumococcal autolysin in the process of enzymatic activation (conversion). Gene. **61**: 13-19.
- Sánchez-Puelles, J.M., Ronda, C., García, J.L., García, P., López, R. y García, E. 1986b. Searching for autolysin functions. Eur. J. Biochem. **158**: 289-293.
- Sanger, F., Nicklen, S y Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74**: 5463-5467.
- Sanz, J M., Díaz, E. y García, J. L. 1992. Studies on the structure and function of the N-terminal domain of the pneumococcal murein hydrolases. Mol. Microbiol. **6**: 921-931.
- Sanz, J.M., López, R. y García, J.L. 1988. Structural requirements of choline derivatives for "conversion" of pneumococcal amidase. FEBS Lett. **232**: 308-312.
- Sato, Y., Cowell, J.L., Sato, H., Burstyn, D.G. y Manclark, C.R. 1983. Separation and purification of the hemoagglutinins from *Bordetella pertussis*. Infect. Immunol. **41**: 313-320.
- Saxena, I. M., Brown, R. M., Jr., Fevre, M., Geremia, R. A. y Herissat, B. 1995. Multidomain architecture of -glycosyl transferases: Implications for mechanism of action. J. Bacteriol. **177**: 1419-1424.
- Segrest, J.P., Jackson, R.L., Morrisett, J.D. y Gotto, A.M. 1974. A molecular theory of lipid-protein interactions in the plasma lipoproteins. FEBS. Lett. **38**: 247-253.
- Sekine Y. y Ohtsubo, E. 1989. Frameshifting is required for production of the transposase encoded by insertion sequence 1. Proc. Natl. Acad. Sci. USA **86**: 4609-4613.
- Shan, F. 1990. Pneumococcus and influenza. Lancet. **335**: 898-901.
- Shiroza, T., Ueda, S. y Kuramitsu, H. K. 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. J. Bacteriol. **169**: 4263-4270.
- Silen, J.L., McGrath, C.N., Smith, K.R. y Agard, D.A. 1988. Molecular analysis of the gene encoding α -lytic protease: evidence for a preproenzyme. Gene **69**: 237-244.
- Simonen, M. y Palva, I. 1993. Protein secretion in *Bacillus* species. Microbiol. Rev. **57**: 109-137.
- Smith, D. K., Kassam, T., Singh, B. y Elliot, J. F. 1992. *Escherichia coli* has two homologous glutamate decarboxylase genes that map to distinct loci. J. Bacteriol. **174**: 5820-5826.
- Smith, H. E., Reek, F. H., Vecht, U., Gielkens, A. L. J. y Smits, M. A. 1993. Repeats in an extracellular protein of weakly pathogenic strains of *Streptococcus suis* type 2 are absent in pathogenic strains. Infect. Immun. **61**:3318-3326.
- Soucaille, P. y Goma, G. 1986. Acetonobutylic fermentation by *Clostridium acetobutylicum* ATCC824: autobacteriocin production, properties and effects. Curr. Microbiol. **13**: 163-169.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments by gel electrophoresis. J. Mol. Biol. **98**: 507-517.
- Sparrow, J.T. y Gotto, A.M., Jr. 1982. Crit. Rev. Biochem. **13**: 87-107.

- Stenman, S y Vaheri, J. 1978. Distribution of a major connective tissue protein, fibronectin, in normal human tissues. *J. Exp. Med.* 147: 1054-1064.
- Studier, F.W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219: 37-44.
- Swank, R. T. y Munkres, K.D. 1971. Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Anal. Biochem.* 39: 462-477.
- Takahashi, N., Takahashi, Y. y Putnam, F.W. 1985. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine rich α_2 -glycoprotein of human serum. *Proc. Natl. Acad. Sci. USA.* 82: 1906-1910.
- Takumi, K. y Kawata, T. 1970. *Japan. J. Microbiol.* 14: 57
- Talay, S.R., Valentin-Weigand, P., Timmis, K.N. y Chatwal, G.S. 1994. Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of *Streptococcus pyogenes*. *Mol. Microbiol.* 13: 531-539
- Taylor, N.S., Thorne, G.M y Bartlett, J.G. 1981. Comparison of two toxins produced by *Clostridium difficile*. *Infect. Immun.* 34: 1036-1043.
- Titani, K., Takio, K., Handa, M. y Ruggeri, Z.M. 1987. Amino acid sequence of the von Willebrand factor-binding domain of platelet membrane glycoprotein. *Proc. Natl. Acad. Sci.* 84: 5610-5614.
- Tomasz, A. 1967. Choline in the cell wall of a bacterium: novel type of polymer-linked choline in pneumococcus. *Science* 157: 694-697.
- Tomasz, A. 1968. Biological consequences of the replacement of choline by ethanolamine in the cell wall of pneumococcus: chain formation, loss of transformability, and loss of autolysis. *Microbiology*, 59: 86-93.
- Tomasz, A. 1970. Requirement for protein synthesis during induction of competence. *J. Bacteriol.* 101: 860-871.
- Tomasz, A. 1979. From penicillin binding proteins to the lysis and death of bacteria: A 1979 view. *Rev. Infect. Dis.* 1: 434-467.
- Tomasz, A. 1981. Surface components of *Streptococcus pneumoniae*. *Rev. Infect. Dis.* 3: 190-211.
- Tomasz, A. y Hochtiss, R.D. 1964. Regulation of the transformability of pneumococcal autolytic enzyme by inhibitors of cell wall synthesis. *Proc. Natl. Acad. Sci. USA.* 72: 4162-4166.
- Tomasz, A. y Waks, S. 1975. Mechanism of action of penicillin: triggering of pneumococcal autolytic enzyme by inhibitors of cell wall synthesis. *Proc. Natl. Acad. Sci. USA.* 72: 4162-4166.
- Tomasz, A. y Westphal, M. 1971. Abnormal autolytic enzyme in pneumococcus with altered teichoic acid composition. *Proc. Natl. Acad. Sci. USA* 68: 2627-2630.
- Tomasz, A., Albino, A. y Zanati, E. 1970. Multiple antibiotic resistance in a bacterium with suppressed autolytic system. *Nature* 227: 138-140.
- Tomasz, A., Westphal, M., Briles, E. B. y Fletcher, P. 1975. *J. Supramol. Struct.* 3: 1-16.
- Ueda, S., Shiroza, T. y Kuramitsu, H.K. 1988. Sequence analysis of the *gtfC* gene from *Streptococcus mutans* GS-5. *Gene* 69: 101-109.
- Van der Westhuizen, A., Jones, D. T. y Woods, D. R. 1982. Autolytic activity and butanol tolerance of *Clostridium acetobutylicum*. *Appl. Env. Microbiol.* 44: 1277-1282.
- Van Vliet, F., Cunin, R., Jacobs, A., Piette, J., Gigot, D., Lauwereys, M., Pierard, A. y N. Glansdorff. 1984. Evolutionary divergence of genes for ornithine and aspartate

- carbamoyl-transferases; complete sequence and mode of regulation of the *Escherichia coli* *argF* gene: comparison of *argF* with *argI* and *pyrB*. *Nucleic Acids Res.* **12**:6277-6289.
- Vieira, J., y Messing, J. 1982. The pUC plasmids a M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, **19**: 259-268.
- Von Eichel-Streiber, C. y Sauerborn, M. 1990. *Clostridium difficile* toxins A and B. *Mol. Gen. Genet.* **233**: 260-268.
- Von Eichel-Streiber, C., Laufenberg-Feldmann, R., Saringen, S., Schulze, J. y Sauerborn, M. 1992a. Comparative sequence analysis of the *Clostridium difficile* A and B. *Mol Gen Genet* **233**: 260-268.
- Von Eichel-Streiber, C., Sauerborn, M. y Kuramitsu, H.K. 1992b. Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. *J. Bacteriol.* **174**: 6707-6710.
- Walter, K. A., Bennett, G.N. y Papoutsakis, E. T. 1992. Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isoenzymes genes. *J. Bacteriol.* **174**: 7149-7158.
- Waltman, W. D., McDaniel, H. L. S. y Briles, E. 1990. Variation in the molecular weight of PspA (pneumococcal surface protein A) among *Streptococcus pneumoniae*. *Microb. Pathog.* **8**: 61-69.
- Webster, J. R., Reid, S. J., Jones, D. T. y Woods, D.R. 1981. Purification and characterization of an autolysin from *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **41**: 371-374.
- Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F. y Gesteland, R.F. 1989. *Escherichia coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.* **1**:159-169.
- Wernars, K., Heuvelman, K., Notermans, S. Domman, E., Leimeister-Wachter, M. y Chakraborty, T. 1992. Suitability of the *prfA* gene, which encodes regulator of virulence genes in *Listeria monocytogenes*, in the identification of pathogenic *Listeria* spp. *Appl. Environ. Microbiol.* **58**: 765-768.
- Wikstrom, M.B., Dahlen, G., Kaijser, B. y Nygen, H. 1984. *Infect. Immun.* **44**: 33-37.
- Wilkinson, S. R. y Young, M. 1993. Wide diversity of genome size among different strains of *Clostridium acetobutylicum*. *J. Gen. Microbiol.* **139**: 1069-1076.
- Wilkinson, S. R., Young, M. 1995. Physical map of the *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052 chromosome. *J. Bacteriol.* **177**: 439-448.
- Wilkinson, S. R., Young, M., Goodacre, R., Morris, J.G., Farrow, J.A.E. y Collins, M.D. 1995. Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum*. *FEMS Microbiol Lett.* **125**: 199-204.
- Wolley, R.C. y Morris, J.G. 1990. Stability of solvent production by *Clostridium acetobutylicum* in continuous culture: strain differences. *J. Appl. Bacteriol.* **69**: 718-728.
- Wren, B. W. 1991. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol. Microbiol.* **5**: 797-803.
- Wren, B.W., Clayton, C.L. y Tabaqchali, S. 1990. Nucleotide sequence of *Clostridium difficile* toxin A gene fragment and detection of toxigenic strains by polymerase chain reaction. *FEMS Microbiol. Lett.* **70**: 1-6.
- Yamashita, Y., Hanada, N. y Takehara, T. 1989. Purification of a fourth glucosyltransferase from *Streptococcus sobrinus*. *J. Bacteriol.* **171**: 6265-6270.

- Yanisch-Perron, C., Vieira, J. y Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. *Gene* 33: 103-119.
- Yoshikawa, K., Tsuzuki, H., Fujiwara, T., Nakamura, E., Iwamoto, H., Matsumoto, K., Shin, M. y Yoshida, N. 1992. Purification, characterization and gene cloning of a novel glutamic acid-specific endopeptidase from *Staphylococcus aureus* ATCC 12600. *Biochim. Biophys. Acta.* 1121: 221-228.
- Yoshino, S., Ogata, S. y Hayashida, S. 1982. Some properties of autolysin of *Clostridium sacharoperbutylaceticum*. *Agr. Biol. Chem.* 46: 1243-1248.
- Yother, J. y Briles, D.E. 1992a. Structural properties and evolutionary relationships of PspA, a surface protein of *Streptococcus pneumoniae*, as revealed by sequence analysis. *J. Bacteriol.* 174: 601-609.
- Yother, J. y White, J.M. 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. *J. Bacteriol.* 176: 2976-2985.
- Yother, J., Handsome, G.L. y Briles, D.E. 1992b. Truncated forms of PspA that are secreted from *Streptococcus pneumoniae* and their use in functional studies and cloning of the *pspA* Gene. *J. Bacteriol.* 174: 610-618.
- Young, M., Minton, N. P. y Staudenbauer, W. L. 1989. Recent advances in the genetic of clostridia. *FEMS Microbiol. Rev.* 63:301-326.
- Zhou, L., Hui, F.M. y Morrison, D.A. 1995. Characterization of IS1167, a new insertion sequence in *Streptococcus pneumoniae*. *Plasmid* 33: 127-138.