

DISEÑO DE OLIGONUCLEÓTIDOS PARA EL ESTUDIO DE GENES CELULOLÍTICOS Y SOLVENTOGÉNICOS EN CEPAS COLOMBIANAS DE *Clostridium* sp. (CLOSTRIDIACEAE)

Oligonucleotide Probe Design for the Study of Cellulolytic and Solventogenic Genes in Colombian *Clostridium* sp. Strains (Clostridiaceae)

JOSÉ DAVID MONTOYA SOLANO¹, Químico Farmacéutico; ZULMA ROCÍO SUÁREZ MORENO², M.Sc.; DIEGO MAURICIO RIAÑO PACHÓN³, Biólogo; DOLLY MONTOYA CASTAÑO¹, M.Sc., Dr. rer. nat.; FABIO ANCÍZAR ARISTIZÁBAL GUTIÉRREZ⁴, Ph. D.

¹Grupo de Bioprocesos y Bioprospección, Instituto de Biotecnología, Universidad Nacional de Colombia, Sede Bogotá, ciudad universitaria, Carrera 30 No. 45-03. AA. 14490. Teléfono: 57-1 316 50 00, exts. 16971 - 16954. jdmontoyas@unal.edu.co - dmontoyac@unal.edu.co

²Grupo de Bacteriología y Bacteriología Vegetal, Laboratorios ICGEB Trieste, Italia. Teléfono: +39-40-375 73 17. suarez@icgeb.org

³Departamento de Biología Molecular, Universidad de Potsdam. Potsdam, Alemania. Teléfono: +49-331-977 28 09.

diriano@rz.uni-potsdam.de

⁴Departamento de Farmacia, Universidad Nacional de Colombia, Sede Bogotá, ciudad universitaria, Carrera 30 No. 45-03. AA. 14490. Teléfono: 57-1 316 50 00, ext. 14643. faaristizabal@unal.edu.co

Presentado el 2 de octubre de 2006, aceptado el 16 de mayo de 2007, correcciones 3 de octubre de 2007.

RESUMEN

El objetivo de este estudio fue analizar las rutas metabólicas para la producción de solventes y degradación de celulosa en cepas colombianas promisorias del género *Clostridium*. Para ello se diseñaron sondas de hibridación que sirvieran para posteriores estudios de mejoramiento genético de las cepas. Se construyó la base de datos denominada MULTI-CLOST en Microsoft Access® con las secuencias de 485 genes involucrados en las rutas metabólicas arriba mencionadas, provenientes de 45 especies bacterianas y 10 especies fúngicas. Los genes fueron agrupados de acuerdo al tipo de enzima y a los dominios catalíticos o de unión a sustrato en el caso de las celulasas. Cada grupo se sometió a alineamiento múltiple en ClustalW 1.83 y con base en los resultados se crearon subgrupos de similitud mayor al 50%. Se localizaron secuencias conservadas de longitud mayor a 19 nucleótidos en GeneDoc 2.6.002 y sus valores termodinámicos fueron estimados con GeneRunner v3.05, mientras que la sensibilidad y especificidad fue verificada por búsquedas en GenBank usando BLASTN 2.2.8. En total se obtuvieron 94 secuencias conservadas con las siguientes características: longitud promedio de 24

nucleótidos, Tm promedio de 65,8 °C y contenido de (G+C) entre 14,3 y 60,0%. Se determinó que ninguna de las sondas diseñadas forma estructuras secundarias estables con Tm superior a 36,1 °C. De acuerdo a sus características y valores termodinámicos, todas las sondas podrían ser utilizadas en la construcción de un microarreglo o en reacciones de PCR para la identificación de regiones relevantes en el mejoramiento del proceso por ingeniería metabólica.

Palabras clave: sondas de oligonucleótidos, *Clostridium* sp., Microarreglos, Bioinformática.

ABSTRACT

The goal of the present study was to analyze the metabolic pathways involved in solvent production and cellulose consumption by promising Colombian native strains of the genus *Clostridium*. Therefore a set of oligonucleotide probes was designed, with the aim of analyzing potential targets for genetic improvement of the Colombian strains. The database named MULTI-CLOST was created in Microsoft Access® using the sequences from 485 genes involved in solventogenesis, 1,3-propanediol production and cellulolysis from 45 bacterial and 10 fungal species. The genes were grouped according to their respective enzyme function and to the catalytic domain or the substrate binding domain in the case of cellulases. ClustalW 1.83 was used for multiple alignment of every group. Subgroups of sequences with more than 50% identity among themselves were created. Conserved sequences longer than 19 nucleotides were identified using GeneDoc 2.6.002 and their thermodynamic values were calculated with GeneRunner v3.05, while their sensitivity and specificity were verified by searching in GenBank with BLASTN 2.2.8. Ninety-four conserved sequences were obtained with an average 24-nucleotide length, 65.8°C average Tm and a (G+C) content between 14.3% and 60.0%. None of these probes forms stable secondary structures at temperatures higher than 36.1°C. According to the former results, all of the probes could be used in an oligonucleotide microarray or in PCR reactions for the identification of metabolic targets for improvement of the industrial process.

Key words: Oligonucleotide probes, *Clostridium* sp., Microarrays, Bioinformatics.

INTRODUCCIÓN

Este trabajo parte de la necesidad de una herramienta molecular adecuada para la detección de genes involucrados en los procesos de solventogénesis y celulolisis en cepas colombianas promisorias del género *Clostridium*. Dicho género incluye especies solventogénicas y no patógenas empleadas extensamente desde principios del siglo XX para la producción de acetona, butanol y etanol (Jones *et al.*, 1986). Algunas de estas especies tienen la capacidad de utilizar sustratos celulósicos para la fermentación acetobutílica, los cuales reducen significativamente el costo total del proceso. Tradicionalmente la rentabilidad industrial de la biosíntesis de estos solventes y la degradación de celulosa depende de los rendimientos de la cepa y del costo de la fuente de

carbono (Lynd *et al.*, 2002; Pachauri *et al.*, 2006). En el ámbito internacional se han patentado varios procesos consolidados para la producción biotecnológica de solventes con cepas bacterianas transformadas, alcanzando altos rendimientos por medio del uso de sustratos económicos y enzimas de alta eficiencia (Nakamura *et al.*, 2003; Zverlov *et al.*, 2006).

El Instituto de Biotecnología de la Universidad Nacional de Colombia (IBUN) ha trabajado en el empleo de residuos agroindustriales y en la búsqueda de cepas nativas con rendimientos superiores al de las cepas patrón. En la década de 1990 se aislaron 178 cepas de *Clostridium* a partir de suelos colombianos con el ánimo de encontrar cepas hiperproductoras de solventes. En uno de los estudios realizados se seleccionaron 13 cepas teniendo como base la concentración de solventes totales -acetona, butanol y etanol- a partir de glucosa (Montoya *et al.*, 2000). Estudios posteriores evidenciaron que las cepas nativas de *Clostridium* pueden degradar sustratos celulósicos como pectina, xilano, carboximetil-celulosa y celulosa microcristalina (Montoya *et al.*, 2001), mientras que cinco de las cepas producen 1,3-propanodiol (1,3-PD) a partir de glicerol en mayor concentración que *Clostridium butyricum* DSM2478 (Cárdenas *et al.*, 2006). Adicionalmente los análisis de variabilidad molecular por AFLP -*Amplified Fragment Length Polymorfism*-, PFGE -*Pulsed Field Gel Electrophoresis*- e hibridación DNA-DNA sugieren que 10 de las cepas promisorias corresponderían a una nueva especie del género *Clostridium* filogenéticamente muy cercana a *C. butyricum* (Arévalo, 2001; Suárez, 2004; Jaimes *et al.*, 2006). Este conjunto de resultados permite considerar a las cepas nativas como candidatas para el diseño de bioprocesos innovadores de producción de solventes a nivel industrial. Los rendimientos de las cepas promisorias podrían ser mejorados por ingeniería metabólica, lo cual requiere del conocimiento de los genes de las rutas metabólicas involucradas en los procesos de solventogénesis y celulolisis. Los primeros estudios genéticos de las cepas promisorias incluyen la construcción de una librería genómica a partir de la cepa IBUN 22A, en la cual se encontraron ocho clones con actividades exoglucanasa, endoglucanasa y -glucosidasa (Vargas *et al.*, 2002). Estudios posteriores permitieron predecir genes homólogos a *dhaB1* y *dhaT* de *C. butyricum* en las cepas nativas IBUN 22A, IBUN 13A e IBUN 158B; dichos genes codifican para las proteínas Glicerol Deshidratasa y 1,3-Propanodiol Óxidoreductasa involucradas en la producción de 1,3-PD a partir de glicerol (Montoya *et al.*, 2006; Quilaguy *et al.*, comunicación personal).

Un estudio completo de los genes involucrados en las rutas metabólicas de interés requiere de una estrategia más eficiente, basada en la información genómica que ha sido obtenida a nivel mundial para el género *Clostridium*. Sin embargo, hasta la fecha *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium kluyveri* y *Clostridium thermocellum* son las únicas especies solventogénicas y/o celulolíticas no patógenas del género cuyo genoma ha sido completamente secuenciado (Nölling *et al.*, 2001; Copeland *et al.*, 2007a; Copeland *et al.*, 2007b; Seedorf *et al.*, 2007). Un ensayo de hibridación basado en sondas universales constituye una buena alternativa, puesto que permitiría analizar la presencia y expresión de los genes involucrados en todas las rutas metabólicas sin necesidad de conocer previamente la secuencia genómica

exacta de las cepas promisorias. En este trabajo se plantea el uso de un microarreglo como ensayo de hibridación, puesto que los microarreglos poseen alta capacidad de procesamiento de muestras y su costo es comparable con el de otras técnicas para el análisis de perfiles de expresión como SAGE -*Serial Analysis of Gene Expresión*-, SSH -*Suppression Subtractive Hybridization*- y MPSS -*Massively Parallel Signature Sequencing*- (Velvescu *et al.*, 1995; Bunney *et al.*, 2003; Stoughton, 2005). Los microarreglos de DNA son ensayos dirigidos a la secuencia genómica de un organismo que permiten analizar la concentración relativa de una gran variedad de moléculas específicas de DNA o RNA simultáneamente. Las muestras analizadas usualmente corresponden a DNA o RNA total extraído del cultivo bajo las condiciones de estudio, el cuál es hibridado con las sondas del microarreglo para determinar la presencia y nivel de expresión de genes específicos (Stoughton, 2005). Por medio de los microarreglos se pueden identificar fenotipos bacterianos específicos, asignarle función a genes desconocidos, agrupar genes en rutas metabólicas de interés, identificar los genes reguladores de dichas rutas y otras funciones que resultan útiles en la comprensión del metabolismo fermentativo de las cepas nativas de *Clostridium* sp. Por ejemplo, estudios previos de cepas mutadas de *C. acetobutylicum* usando un microarreglo genómico han demostrado la presencia de genes reguladores necesarios para la solventogénesis y la esporulación en el megaplásmido pSOL1 (Tomas *et al.*, 2003; Alsaker *et al.*, 2005). El objetivo de este artículo es proponer una metodología para el diseño de sondas de DNA útiles en la construcción de un microarreglo de oligonucleótidos o en otros ensayos de hibridación enfocados al estudio y mejoramiento de las rutas metabólicas anteriormente mencionadas. Las sondas para microarreglos pueden corresponder a cDNAs preidentificados (RNA total retrotranscrito) o bien a una colección de oligonucleótidos diseñados para genes específicos. El uso de un microarreglo de oligonucleótidos supone ventajas en nuestro caso, puesto que los oligonucleótidos basados en secuencias consenso permiten detectar la expresión de genes conservados en especies cuyo genoma aún no ha sido secuenciado, como es el caso de las cepas colombianas de *Clostridium* sp. La utilidad de esta metodología ha sido demostrada en trabajos como la patente de Graham y Wollweber (1990), quienes diseñaron oligonucleótidos consenso para la detección de α -amilasas en diferentes especies del género *Bacillus*, o bien en el trabajo de Matveeva *et al.* (2004), quienes diseñaron oligonucleótidos consenso dirigidos al genoma del VIH-1 a partir de regiones conservadas en los alineamientos múltiples de sus variantes. El uso de micro-arreglos de oligonucleótidos también ha sido reportado para el género *Clostridium* por Paredes *et al.* (2007), quienes diseñaron un microarreglo con oligómeros de 60 nucleótidos para el análisis metabólico en cepas mutadas de *C. acetobutylicum*. Este artículo presenta una metodología para el diseño de oligonucleótidos a partir de secuencias genéticas publicadas en GenBank para especies bacterianas y fúngicas representativas de las rutas metabólicas de interés, con la cual se busca compensar la baja cantidad de secuencias publicadas hasta la fecha para especies del género *Clostridium*. El diseño está basado en la identificación de secuencias conservadas en la mayor cantidad de especies posible, de forma que cada sonda resultante sea útil para detectar la presencia del mismo gen tanto en dichas especies como en las cepas nativas de *Clostridium* sp. (partiendo de su cercanía evolutiva y/o metabólica). En la metodología propuesta se

han tenido en cuenta aspectos generales del diseño de oligonucleótidos para microarreglos, tales como su especificidad con respecto al gen objetivo, su incapacidad para formar homodímeros o heterodímeros con otras sondas, la ausencia de estructuras secundarias y el control de parámetros como longitud y Tm (punto de fusión) dentro de rangos previamente establecidos (Tolstrup *et al.*, 2003).

MATERIALES Y MÉTODOS

CONSTRUCCIÓN DE LA BASE DE DATOS LOCAL MULTI-CLOST

A partir de una revisión bibliográfica se identificaron las enzimas que participan en la producción de acetato, butirato, etanol, acetona y butanol a partir de piruvato (Fig. 1) así como en la producción de 1,3-propanodiol y dihidroxiacetona-P a partir de glicerol (Fig. 2) en clostridios solventogénicos no patógenos. También se identificaron las enzimas involucradas en la degradación de celulosa hasta glucosa, incluyendo las proteínas estructurales del celulosoma (complejo de degradación de sustratos celulósicos), y en la degradación de polisacáridos de la hemicelulosa hasta los monosacáridos correspondientes (Fig. 3; Schwarz, 2001). Las secuencias de los genes correspondientes fueron obtenidas en GenBank (Benson *et al.*, 2007) y UniProt (Bairoch *et al.*, 2005) para 16 especies del género *Clostridium* y 29 géneros de microorganismos relacionados (Tabla 1). Los registros de dichas secuencias fueron almacenados y categorizados en una base de datos construida en Microsoft Access XP. Para cada registro se recopiló la siguiente información: nombre del gen, número de acceso de GenBank, nombre de la proteína codificada, número de acceso de Uniprot, especie, conglomerado (si el gen pertenece a uno), longitud del gen y longitud de la proteína.

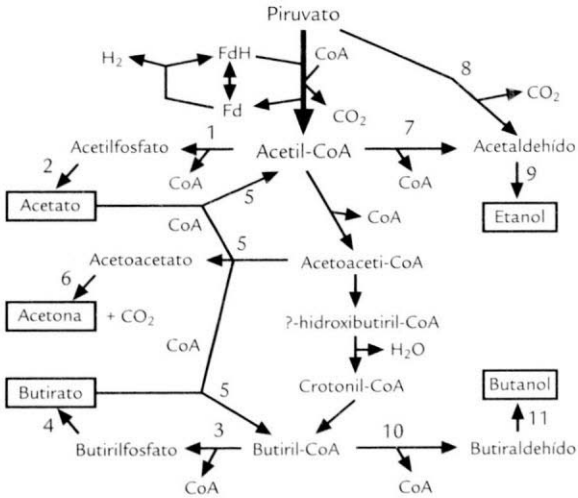


Figura 1. Enzimas involucradas en la producción de solventes de *Clostridium* incluidas en la base de datos. 1. Fosfotransacetilasa 2. Acetato kinasa 3. Fosfotransbutirilasa. 4. Butirato kinasa. 5. Acetoacetil:acil CoA transferasa. 6. Acetoacetato descarboxilasa 7. Acetaldehído deshidrogenasa. 8. Piruvato descarboxilasa. 9. Alcohol deshidrogenasa. 10. Butiraldehído deshidrogenasa. 11. Butanol deshidrogenasa.

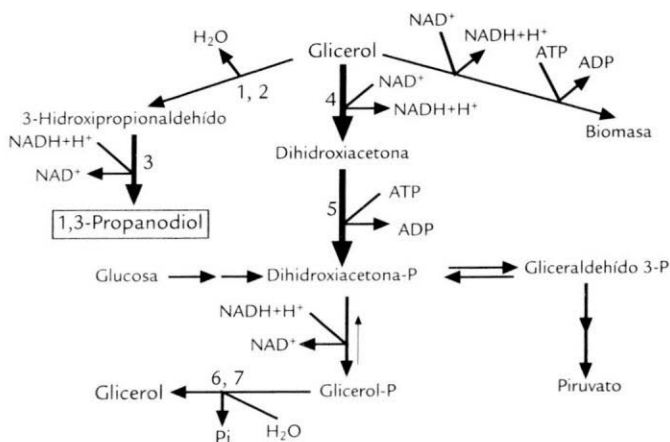


Figura 2. Ruta de asimilación de glicerol en *Clostridium butyricum*. 1. Glicerol deshidratasa. 2. Activador de la glicerol deshidratasa. 3. 1,3-PD deshidrogenasa. 4. Glicerol deshidrogenasa. 5. Dihidroxiacetona kinasa. 6. Glicerol-3-fosfatasa. 7. Glicerol kinasa.

Para esta base de datos local se establecieron las siguientes categorías: (1) ‘Solventogénesis’, con las enzimas fosfotranscarboxilasas, carboxilato kinasas, CoA transferasas, acetoacetato descarboxilasas, aldehído deshidrogenasas y alcohol deshidrogenasas, necesarias para la producción de acetato, butirato, etanol, butanol y acetona (Jones y Woods, 1986; Biebl *et al.*, 1999; Nölling *et al.*, 2001); (2) ‘1,3-Propanodiol’, donde se almacenaron registros para las enzimas glicerol deshidratasa, 1,3-PD deshidrogenasa, glicerol deshidrogenasa, dihidroxiacetona kinasa y glicerol-3-fosfatasa, necesarias para la producción de 1,3-propanodiol y glicerol (Biebl *et al.*, 1999); (3) ‘Celulasas’, donde se incluyeron las proteínas estructurales que forman el andamio y el sistema de anclaje del celulosoma (Bayer *et al.*, 1998), así como celulasas entre las que se cuentan exoglucanasas, endoglucanasas, β -glucosidasas, xilanasas, xilosidasas y mananasas (Coughlan y Mayer, 1992; Lucas *et al.*, 2001). La subcategoría ‘Celulasas’ fue a su vez separada por familias de glicosil hidrolasas, dominios de unión a celulosa y dominios de homología con la capa S considerando la composición multidominio de estas enzimas (Rabinovich *et al.*, 2002). Adicionalmente se decidió incluir la categoría ‘Factor de transcripción Spo0A’, dado que esta proteína de esporulación ha sido reconocida como factor de transcripción de los operones *sol*, *adc* y *bdh* de producción de solventes (Harris *et al.*, 2002). La lista completa de categorías y subcategorías se presenta en la Tabla 2.

CONSTRUCCIÓN DE ALIMENTOS MÚLTIPLES CON ALTA CONSERVACIÓN DE SECUENCIAS

Para cada tipo de enzima las secuencias de los genes registrados en MULTI-CLOST fueron recuperadas a partir de GenBank, tomando siempre la cadena positiva de cada gen. En el caso de la categoría ‘Celulasas’ los grupos fueron formados por familias de dominios catalíticos o de unión a sustrato dentro de cada tipo de enzima (Bayer *et al.*, 1998). Las secuencias de cada grupo fueron alineadas con el programa de alineamiento múltiple ClustalW 1.83 (Thompson *et al.*, 1994) usando la matriz de respectivamente. La calidad de los alineamientos múltiples fue mejorada por eliminación de

Especie	Solventogénesis	Celulasas	Categoría	
			Factor Spo0A	1,3-Propanodiol
<i>Acetivibrio cellulolyticus</i>	-	3	-	-
<i>Aquifex aeolicus</i>	-	-	-	2
<i>Bacillus anthracis</i>	-	-	1	-
<i>Bacillus cereus</i>	-	-	1	4
<i>Bacillus halodurans</i>	12	-	1	3
<i>Bacillus sphaericus</i>	-	-	1	-
<i>Bacillus subtilis</i>	17	-	1	-
<i>Bacteroides cellulosolvens</i>	-	1	-	-
<i>Bacteroides thetaiotaomicron</i>	-	19	-	-
<i>Brucella melitensis</i>	-	-	-	2
<i>Butyrivibrio fibrisolvens</i>	-	7	-	-
<i>Caldicellulosiruptor</i> spp.	-	18	-	-
<i>Cellulomonas fimi</i>	-	12	-	-
<i>Citrobacter freundii</i>	-	-	-	8
<i>Clostridium acetobutylicum</i>	17	32	1	2
<i>Clostridium beijerinckii</i>	9	-	1	1
<i>Clostridium butyricum</i>	-	-	-	8
<i>Clostridium cellulolyticum</i>	-	13	1	-
<i>Clostridium cellulovorans</i>	-	15	-	-
<i>Clostridium difficile</i>	-	-	1	-
<i>Clostridium innocuum</i>	-	-	1	-
<i>Clostridium josui</i>	-	5	-	-
<i>Clostridium longisporum</i>	-	2	-	-
<i>Clostridium pasteurianum</i>	-	-	1	4
<i>Clostridium perfringens</i>	13	-	-	8
<i>Clostridium saccharobutylicum</i>	1	2	-	-
<i>Clostridium saccharoperbutylacetonicum</i>	4	-	-	-
<i>Clostridium stercorarium</i>	-	11	-	-
<i>Clostridium tetani</i>	13	-	1	6
<i>Clostridium thermocellum</i>	2	39	-	-
<i>Emericella nidulans</i>	-	-	-	1
<i>Fibrobacter succinogenes</i>	-	16	-	-
<i>Geobacillus stearothermophilus</i>	-	-	1	2
<i>Hypocrea jecorina</i>	-	21	-	-
<i>Klebsiella pneumoniae</i>	-	-	-	6
<i>Lactobacillus plantarum</i>	-	-	-	6
<i>Mesorhizobium loti</i>	-	-	-	5
<i>Oceanobacillus iheyensis</i>	-	-	1	4
<i>Orpinomyces</i> spp.	-	16	-	-
<i>Piromyces</i> spp.	-	24	-	-
<i>Pseudomonas putida</i>	-	-	-	3
<i>Ruminococcus albus</i>	-	12	-	-
<i>Ruminococcus flavefaciens</i>	-	13	-	-
<i>Saccharomyces cerevisiae</i>	-	-	-	5
<i>Sarcina ventriculi</i>	1	-	-	-
<i>Schizosaccharomyces pombe</i>	-	-	-	3
<i>Sinorhizobium meliloti</i>	-	-	-	6
<i>Thermoanaerobacterium thermosaccharolyticum</i>	2	-	-	-
<i>Thermotoga maritima</i>	-	-	-	3
<i>Zigosaccharomyces rouxii</i>	-	-	-	3
<i>Zymomonas mobilis</i>	4	-	-	-
TOTAL	95	281	14	95

Tabla 1. Registros incluidos por cada especie en las categorías de MULTI-CLOST.

los genes con secuencia menos conservada, hasta alcanzar un valor alineamiento ClustalW1.6 (peso de 1 para las concordancias y 0 en los demás casos; Aiyar, 2000) y los valores de penalización por apertura y extensión de gaps 15 y 6,66 superior a 65 en el promedio del Puntaje de Alineamientos Pareados (PAP) calculado por ClustalW entre todos los genes de cada alineamiento (este valor mínimo se redujo a 50 en las familias enzimáticas con alta variabilidad de secuencia, como es el caso de las celulasas multidominio). Los alineamientos múltiples definitivos fueron construidos con los genes restantes, utilizando los valores de penalización 15 y 6,66 para apertura y extensión de gaps (11 y 1 para las familias enzimáticas con alta variabilidad de secuencia).

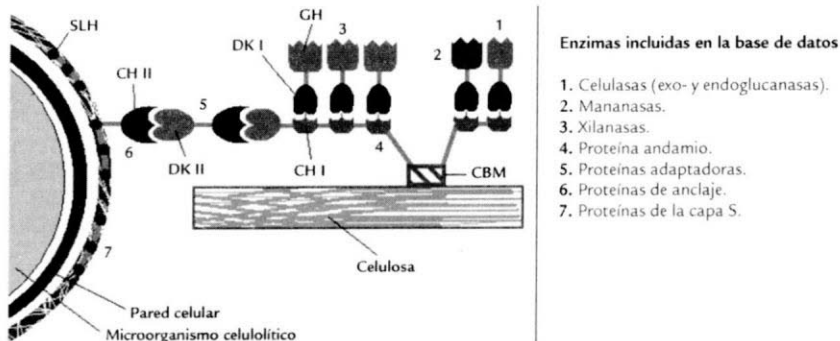


Figura 3. Componentes estructurales y catalíticos del celulosoma, unidos a una célula del microorganismo que lo produce y al sustrato celulósico simultáneamente. Dominios funcionales: SLH, homología con la capa S; CH I y CH II, Cohesin I y II; DK I y DK II, Dockerin I y II; CBM, módulo de unión a celulosa; GH, Glicosil hidrolasa.

DISEÑO DE LAS SONDAS DE OLIGONUCLEÓTIDOS

A partir de los alineamientos definitivos se extrajeron secuencias con longitud mínima de 19 nucleótidos conservadas en más del 50% de los genes de cada alineamiento, utilizando para ello el editor de alineamientos GeneDoc 2.6.002 (Nicholas y Nicholas, 2001). La longitud mínima requerida para las sondas fue establecida a partir de los resultados arrojados por BLASTN para sondas de 17 a 20 nucleótidos dirigidas a un mismo gen de Endoglucanasa (Tabla 6). Para aumentar la probabilidad de que los oligonucleótidos diseñados puedan reconocer sus genes objetivo en las cepas promisorias de *Clostridium* sp., se seleccionaron siempre regiones conservadas en al menos una especie de los órdenes Clostridiales o Bacillales. Las propiedades termodinámicas de las sondas diseñadas fueron calculadas con GeneRunner v3.05 (Hastings Software, NY) y se aceptaron solo aquellas cuya Tm según la fórmula termodinámica de SantaLucia (1998) estuviera en el rango de 44 a 90 °C y que no formaran estructuras secundarias con Tm superior a 37 °C. Estos criterios se establecieron considerando las temperaturas de desnaturalización e hibridación recomendadas usualmente para microarreglos (42 °C y 95 °C respectivamente; Hedge *et al.*, 2000; Schuchhardt *et al.*, 2000; Tomas *et al.*, 2003; Tummala *et al.*, 2003; Rybicki, 2005). Para comprobar la sensibilidad y especificidad de los oligonucleótidos diseñados, éstos fueron sometidos a búsquedas en GenBank 141.0 (Benson *et al.*, 2007) usando la aplicación BLASTN 2.2.8 (Altschul *et al.*, 1997). Se seleccionaron aquellos oligonucleótidos para los

cuales todos los registros reportados por BLASTN con Valor E (*Expected Value*) inferior a 0,05 correspondieran a genes con función equivalente a los genes del alineamiento original (los resultados de BLASTN son más significativos cuando su valor E es menor, pero el valor E tiende a ser conservativo cuando la secuencia de búsqueda es muy corta, de manera que para secuencias de longitud menor a 40 nucleótidos los resultados con valor E < 0,05 se pueden considerar específicos; Altschul *et al.*, 1997). Por último se estableció la ubicación de cada sonda de oligonucleótidos en sus genes objetivo usando el programa Artemis v5 (Rutherford *et al.*, 2000) y se determinó el dominio funcional correspondiente a cada una por medio de búsquedas de las proteínas codificadas por los genes objetivo en la base de datos CDD v1.65 (*Conserved Domain Database*) usando RPS-BLAST 2.2.6 (*Reversed Position Specific BLAST*; Altschul *et al.*, 1997).

Acidogénesis y solventogénesis (95 registros)		Celulasas (282 registros)		1,3-Propanodiol (94 registros)	
Subcategoría	# Reg.	Subcategoría	# Reg.	Subcategoría	# Reg.
Fosfotransacetilasa	7	Exoglucanasas	23	Glicerol deshidratasa	17
Acetato kinasa	8	Endoglucanasas	92	Activador de la glicerol deshidratasa	2
Butirato kinasa	6	Mananasas	16	1,3-Propanodiol deshidrogenasa	11
Fosfotransbutirilasa	6	Xilanasas	47	Regulador transcripcional	2
Piruvato descarboxilasa	3	Proteínas andamio	9	Glicerol deshidrogenasa	15
Acetoacetato descarboxilasa	3	Celulasas	22	Dihidroxiacetona kinasa	26
Aldehído deshidrogenasa	14	β -glucanasas	9	Glicerol kinasa	17
Aldehído-alcohol deshidrogenasa	5	β -glucosidasas	29	Glicerol-3-fosfatasa	4
Alcohol deshidrogenasa	18	Xilosidasas	11		
Alcohol deshidrogenasa de cadena corta	3	Arabinosidasas	7		
Butanol deshidrogenasa	13	Celodextrinasas	3		
Acetoacetyl:acil CoA transferasa A	4	Celodextrin fosforilasas	2	Factores de transcripción Spo0A (14 registros)	
Acetoacetyl:acil CoA transferasa B	5	Celobiosa fosforilasas	2	Subcategoría	# Reg.
		Proteínas de la capa S	3	No posee subcategoría	14
		Proteínas adaptadoras del andamio	2		
		Proteínas de anclaje del celulosoma	5		

Tabla 2. Número de registros incluidos en cada categoría y subcategoría de la base de datos MULTI-CLOST.

RESULTADOS

CONSTRUCCIÓN DE LA BASE DE DATOS MULTI-CLOST

La base de datos construida en Microsoft Access XP fue registrada bajo el nombre MULTI-CLOST ante la Dirección Nacional de Derecho de Autor de Colombia (Libro 13, Tomo 17, Partida 142 del 20 de noviembre de 2006; disponible en

<http://www.ibun.unal.edu.co/lineas/bioprosesos/index.htm>). Esta base de datos contiene registros de 485 genes que codifican para las enzimas y proteínas estructurales involucradas en la acidogénesis, solventogénesis, degradación de celulosa y producción de 1,3-PD en 55 especies bacterianas y fúngicas, permitiendo recuperar las secuencias de dichos genes y proteínas en las bases de datos GenBank (Benson *et al.*, 2007) y UniProt (Bairoch *et al.*, 2005). De estos registros 95 corresponden a la categoría 'Solventogénesis', 94 a la categoría '1,3-Propanodiol', 14 a la categoría 'Factor de esporulación Spo0A', y 282 a la categoría 'Celulasas', la más numerosa considerando la subdivisión adicional por familias de dominios funcionales.

Subcategoría	Organismo	Nombre del gen	No. Acceso (GenBank)
Fosfotransacetilasa	<i>Clostridium thermocellum</i>	<i>pta</i>	AF041841
Acetato kinasa	<i>Bacillus subtilis</i>	BSU29470	Z99119
Butirato kinasa	<i>Clostridium beijerinckii</i>	<i>buk</i>	AY463103
Fosfotransbutirilasa	<i>Clostridium acetobutylicum</i>	CAC3076	AE007804
Piruvato descarboxilasa	<i>Zymomonas mobilis</i>	<i>pdc</i>	M15393
Acetoacetato descarboxilasa	<i>C. saccharoperbutylacetonicum</i>	<i>adc</i>	AY251646
Aldehído-alcohol deshidrogenasa	<i>Clostridium tetani</i>	CTC02182	AE015943
Butanol deshidrogenasa	<i>Clostridium perfringens</i>	CPE1256	AP003189
Acetoacetyl:acil CoA transferasa A	<i>Clostridium acetobutylicum</i>	CAP0163	AE001438
Acetoacetyl:acil CoA transferasa B	<i>Bacillus halodurans</i>	BH2258	AP001514
Exoglucanasas	<i>Clostridium thermocellum</i>	<i>celK</i>	AF039030
Endoglucanasas	<i>Hypocrea jecorina</i>	<i>egl4</i>	Y11113
Mananasas	<i>Piromyces</i> sp.	<i>manC</i>	X97520
Xilanasas	<i>Ruminococcus albus</i>	<i>xynC</i>	AB057589
Proteínas andamio	<i>Ruminococcus flavefaciens</i>	<i>scaA</i>	AJ278969
β -glucanasas	<i>Clostridium thermocellum</i>	<i>licB</i>	X63355
β -glucosidasas	<i>Clostridium longisporum</i>	<i>abgA</i>	L49336
Xilosidasas	<i>Clostridium stercorarium</i>	<i>bxiB</i>	AJ508405
Arabinosidasas	<i>Bacteroides thetaiotaomicron</i>	BT1781	AE016933
Celodextrinasas	<i>Butyrivibrio fibrisolvens</i>	<i>ced1</i>	X55732
Proteínas de la capa S	<i>Clostridium thermocellum</i>	<i>slpA</i>	U79117
Glicerol deshidratasa	<i>Klebsiella pneumoniae</i>	<i>dhaB2</i>	U60992
1,3-Propanodiol deshidrogenasa	<i>Clostridium pasteurianum</i>	<i>dhaT</i>	AF006034
Regulador transcripcional	<i>Citrobacter freundii</i>	<i>dhaR</i>	U09771
Glicerol deshidrogenasa	<i>Clostridium butyricum</i>	<i>dhaD1</i>	AY138581
Dihidroxiacetona kinasa	<i>Bacillus cereus</i>	BC0981	AE017001
Glicerol-3-fosfatasa	<i>Saccharomyces cerevisiae</i>	<i>gpp2</i>	D50469
Factor de transcripción Spo0A	<i>Clostridium difficile</i>	<i>spo0A</i>	U43514

Tabla 3. Secuencias representativas de cada subcategoría de MULTI-CLOST con su número de acceso a GenBank.

Las especies mayoritarias de la categoría 'Celulasas' son *C. acetobutylicum* y *C. thermocellum*, las cuales representan el 25% de dicha categoría, mientras que en 'Solventogénesis' el

mayor porcentaje (36%) corresponde a las especies *C. acetobutylicum* y *Bacillus subtilis*. Por su parte, la categoría '1,3-Propanodiol' contiene el 25% de secuencias de las especies *Clostridium butyricum*, *Clostridium perfringens* y *Citrobacter freundii*. En la Tabla 2 se muestra la cantidad de registros que fueron agregados a la base de datos para cada subcategoría, mientras que en la Tabla 3 se enlistan los números de acceso en GenBank para secuencias representativas de cada subcategoría almacenadas en MULTI-CLOST. Los números de acceso de GenBank y UniProt contenidos en cada registro permiten recuperar la secuencia genética e información de conglomerados a partir de GenBank, así como la secuencia proteica e información biológica a partir de UniProt. Todas las enzimas incluidas en cada categoría fueron utilizadas para realizar los alineamientos múltiples.

CONSTRUCCIÓN DE ALINEAMIENTOS MÚLTIPLES CON ALTA CONSERVACIÓN DE SECUENCIAS

Las secuencias de los genes almacenados en MULTI-CLOST fueron usadas para crear los alineamientos a partir de los cuales se diseñarían las sondas de oligonucleótidos. Se establecieron como criterios de selección los valores PAP mínimos 65 y 50 de acuerdo con el contenido de secuencias conservadas en cada alineamiento, eliminándose de cada uno los genes que produjeran PAP inferiores a estos valores. En el caso de las celulasas se usaron parámetros más flexibles (apropiados para enzimas multi-dominio) y sus dominios funcionales se alinearon independientemente. Para los alineamientos de genes de las categorías 'Solventogénesis', '1,3-Propanodiol' y 'Factor de esporulación Spo0A' se usaron las penalidades de apertura y extensión de gaps [15-6,66], establecidas por defecto en ClustalW y apropiadas para genes conservados en toda su extensión a través de diversas especies. En los alineamientos definitivos de celulasas fue necesario flexibilizar dichas penalidades, de forma que se pudieran alinear fácilmente los dominios conservados en cada familia enzimática. En estos casos se utilizan las penalidades de apertura y extensión [11-1], correspondientes a las penalidades por defecto usadas en algunas aplicaciones de BLAST (Altschul *et al.*, 1997). De esta forma los alineamientos definitivos fueron construidos con tres combinaciones de parámetros: 65 - [15- 6,66], 50 - [15- 6,66] y 50 - [11-1] (PAP - [apertura- extensión de gaps]). Con el uso de dichas combinaciones de parámetros se obtuvieron en total 56 alineamientos definitivos para las cuatro categorías de MULTI-CLOST. El número de alineamientos obtenidos en cada categoría con las diferentes combinaciones de parámetros se muestra en la Tabla 4.

Categorías	PAP	Penalidades	No. de alineamientos
Solventogénesis	50	15- 6,66	7
	65	15- 6,66	9
1,3-Propanodiol	50	15- 6,66	7
	65	15- 6,66	3
Factor de esporulación Spo0A	50	15- 6,66	1
	65	15- 6,66	1
Celulasas	50	11- 1	10
		15- 6,66	15
	65	15- 6,66	3

Tabla 4. Alineamientos múltiples agrupados por las categorías de la base de datos MULTI-CLOST y por los criterios de alineamiento.

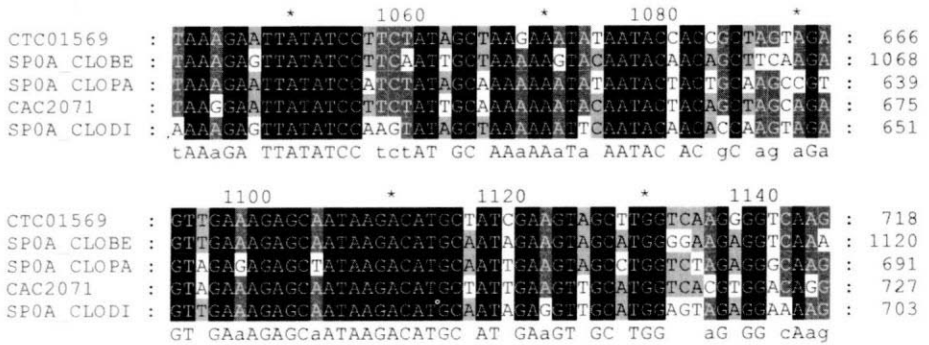


Figura 4. Alineamiento múltiple para el factor de esporulación Spo0A de cinco especies de *Clostridium* usando un PAP mínimo de 65 y penalidades 15- 6,66 (apertura, extensión de gaps). La última línea muestra la secuencia consenso, donde las letras en mayúscula indican posiciones completamente conservadas y las letras en minúscula indican posiciones conservadas en cuatro de las cinco secuencias. Los bloques en negro indican regiones idénticas, mientras que los bloques en diferentes tonos de gris muestran regiones parcialmente conservadas.

Dada la dificultad inherente al alineamiento de genes provenientes de especies tan diversas como las analizadas, el 77% de los alineamientos definitivos contiene dos a cuatro genes y el 18% contiene cinco o seis genes. En la figura 4, a manera de ejemplo, se observa el alineamiento múltiple de genes correspondientes a factores de esporulación Spo0A pertenecientes a varias especies de clostridios.

DISEÑO DE LAS SONDAS DE OLIGONUCLEÓTIDOS

La estrategia de diseño de oligonucleótidos en este trabajo está enfocada a la obtención de sondas que permitan identificar cada gen de interés en cualquiera de las especies que componen los alineamientos, así como en las cepas colombianas de *Clostridium* sp. La longitud mínima de las sondas se estableció a partir de los resultados de búsquedas en GenBank con sondas de longitud creciente diseñadas para un gen Endoglucanasa usando BLASTN. Se encontró que con sondas de longitud inferior a 19 nucleótidos resulta imposible distinguir los resultados específicos de los inespecíficos a partir de los valores de referencia 'Score' y 'Valor E' (Tabla 6; Altschul *et al.*, 1997). Dicha longitud se encuentra en el rango de valores reportados para el diseño de sondas de oligonucleótidos para microarreglos (Guckenberger *et al.*, 2002, Lee *et al.*, 2004). En total se obtuvieron 94 sondas de oligonucleótidos, algunas de las cuales se enlistan en la tabla 5 y cuyos resultados en la búsqueda con BLASTN con Valor E inferior a 0,05 correspondieron en cada caso a las enzimas para las cuales fueron diseñadas (la lista completa de sondas está disponible como material suplementario en <http://www.ibun.unal.edu.co/lineas/bioprosesos/index.htm>). En promedio con cada sonda se pueden reconocer tres genes específicos para la enzima objetivo en GenBank tras las búsquedas correspondientes con BLASTN, lo cual sugiere que existe una buena probabilidad de reconocer los genes homólogos en las cepas promisorias de *Clostridium*. Los resultados de las búsquedas con RPS-BLAST indican que las sondas diseñadas para las rutas de acidogénesis, solventogénesis y producción de 1,3-PD reconocen siempre el mismo dominio en cada gen, mientras que en el caso de las celulasas cada sonda puede identificar varias enzi-

mas de diversa función que poseen el mismo dominio funcional (considerando el carácter multidominio de las celulasas).

Oligo: CAGCTTGAATTCTATCAGTG			Oligo: CAGCTTGAATTCTATCAGT		
Longitud: 20 nucleótidos			Longitud: 19 nucleótidos		
Resultado	Puntaje	valor E	Resultado	Puntaje	valor E
Celulasa de <i>Clostridium</i>	40*	0,009*	Celulasa de <i>Clostridium</i>	38*	0,034*
Celulasa de <i>Clostridium</i>	40*	0,009*	Celulasa de <i>Clostridium</i>	38*	0,034*
Celulasa de <i>Clostridium</i>	40*	0,009*	Celulasa de <i>Clostridium</i>	38*	0,034*
DNA intergénico humano	36	0,13	DNA intergénico humano	36	0,13
DNA intergénico humano	36	0,13	DNA intergénico humano	36	0,13
DNA intergénico de <i>A. thaliana</i>	36	0,13	DNA intergénico de <i>A. thaliana</i>	36	0,13
DNA intergénico humano	34	0,52	DNA intergénico humano	34	0,53
Proteína de unión de rata	34	0,52	Proteína de unión de rata	34	0,53
Factor anemia Fanconi humano	34	0,52			
Receptor de limpieza, <i>T. gondii</i>	34	0,52			
Oligo: CAGCTTGAATTCTATCAG			Oligo: CAGCTTGAATTCTATCA		
Longitud: 18 nucleótidos			Longitud: 17 nucleótidos		
Resultado	Puntaje	valor E	Resultado	Puntaje	valor E
Celulasa de <i>Clostridium</i>	36*	0,13*	Celulasa de <i>Clostridium</i>	34*	0,53*
Celulasa de <i>Clostridium</i>	36*	0,13*	Celulasa de <i>Clostridium</i>	34*	0,53*
Celulasa de <i>Clostridium</i>	36*	0,13*	Celulasa de <i>Clostridium</i>	34*	0,53*
DNA intergénico humano	36*	0,13*	DNA intergénico humano	34*	0,53*
DNA intergénico humano	36*	0,13*	DNA intergénico humano	34*	0,53*
DNA intergénico de <i>A. thaliana</i>	34	0,53	DNA intergénico de <i>A. thaliana</i>	34*	0,53*
DNA intergénico humano	34	0,53	DNA intergénico humano	34*	0,53*
Proteína de unión de rata	34	0,53	Proteína de unión de rata	34*	0,53*

Tabla 6. Resultados de la búsqueda con oligonucleótidos de longitud creciente para Endoglucanasa GH48 en GenBank usando BLASTN 2.2.8. Los resultados rotulados como "Celulasa de *Clostridium*", corresponden a registros de endoglucanasas (resultados específicos), mientras que las cifras marcadas con un asterisco (*), corresponden al mayor puntaje y el menor Valor E obtenidos en cada búsqueda. Se observa que con los oligos de 17 y 18 nucleótidos se obtiene el mayor puntaje y el menor Valor E tanto para los resultados específicos como para algunos inespecíficos. Por ello se estableció como norma para el diseño de sondas una longitud mínima de 19 nucleótidos. Nota: Las magnitudes del mayor Puntaje y el menor Valor E no son comparables entre los cuatro oligos, ya que éstas dependen directa e inversamente de la longitud de la secuencia de búsqueda en BLASTN.

Las sondas diseñadas se encuentran distribuidas de la siguiente forma: 25 corresponden a la categoría 'Solventogénesis', 48 a la categoría 'Celulasas' (con sondas para varias familias de dominios funcionales en cada tipo de enzima), 19 a '1,3-Propanodiol' y dos al factor Spo0A. Los valores característicos promedio de las sondas diseñadas fueron los siguientes: longitud de 24 nucleótidos, Tm termodinámica de 65,8 °C, contenido de (G+C) de 39,0%, Tm de la horquilla más estable (cuando existe) de 10,4 °C y ΔG de unión al objetivo de -37,4 kCal/Mol. De esta forma, se concluye que todas las sondas diseñadas pueden ser usadas para la construcción de un microarreglo hibridizado a 42 °C (Cronin *et al.*, 2001) o bien para un ensayo de Southern Blot realizado en condiciones similares (Harris *et al.*, 2002). Se diseñaron además sondas para control positivo, negativo y de purificación de rRNA utilizando los mismos criterios descritos anteriormente; la sonda de control positivo 5'-

ATCATCCCTAACTCAACTGGTGCTG-3' se diseñó a partir de la gliceraldehído-3P-deshidrogenasa de *C. acetobutylicum*, mientras que la sonda de control negativo 5'-TGACAGATGAATCCGACACATGCAG-3' fue diseñada con base en el gen de la opina, cuyo producto (OPN1SW, GenBank: NM_001708) participa en la transducción de la señal visible en la región azul del espectro electromagnético hacia una señal electroquímica -proceso que ocurre en el ojo humano-. La sonda de control de purificación 5'-TGGTCCGTACAATGAGATGCAACT-3' fue diseñada a partir del gen 16SrRNA de la cepa IBUN 22A (Montoya *et al.*, 2001). Algunos de los oligonucleótidos diseñados también podrían ser utilizados como cebadores para PCR, bien sea en reacciones individuales o en múltiplex PCR. En cada caso uno de los oligonucleótidos diseñados actuaría como cebador forward, mientras que como cebador reverse se podría utilizar un oligonucleótido corto de secuencia aleatoria, como el cebador RAPD-#3 (5'-GTAGACCCGT-3'; Amersham Pharmacia Biotech, Suecia) reportado para RAPD - *Random Amplified Polymorphic DNA*- en *Clostridium difficile* (Peláez *et al.*, 2002). Se seleccionaron 27 oligonucleótidos de los diseñados con la metodología presentada anteriormente, utilizando para ello los criterios de diseño de cebadores de PCR descritos por Rybicki (2005). Estos oligonucleótidos podrían ser utilizados en cuatro reacciones de Múltiplex RT-PCR en Tiempo Real separadas de acuerdo a su temperatura de anillado. El primer grupo (sondas I-3-A, I-6-A, III-3, I-16-B, III-6-A) podría ser empleado en una reacción con una temperatura de anillado de 54 °C; el segundo grupo (sondas I-12-B, I-13, II-11, III-2-A, II-17-A, II-17-B, II-21-A, II-22-C) en una reacción con temperatura de anillado máxima de 58 °C; el tercer grupo (sondas II-1.2, II-7-A, III-1-B, II-12-A, III-2-B, III-5, II-21-B) en una reacción con temperatura de anillado de 63 °C y el cuarto grupo (sondas I-9, II-8-B, II-10-A, II-10-B, II-15.2, III-8-A, III-9) en una reacción con temperatura de anillado de 67 °C. Cabe mencionar que los parámetros termodinámicos predichos con herramientas bioinformáticas deben ser ajustados a partir de los resultados de ensayos de hibridación preliminares.

DISCUSIÓN

La metodología seguida en este trabajo para el diseño de las sondas de oligonucleótidos está examinada a encontrar secuencias conservadas a través de una variedad de microorganismos de los dominios Eubacteria y Eukarya, todos ellos representativos de las rutas metabólicas de interés. Esta estrategia es práctica cuando se recurre a un ensayo de hibridación en el cual la cantidad de sondas es limitada. Sin embargo, en un microarreglo puede incluirse una gran cantidad de sondas, de manera que las secuencias conservadas y sus posibles réplicas dejan muchos espacios libres que pueden ser utilizados para pruebas adicionales. Con el fin de obtener huellas de DNA de cada cepa y dar indicios sobre las relaciones evolutivas de sus genes se podría recurrir a una estrategia adicional: extraer de los alineamientos secuencias con los mismos criterios de longitud, temperatura de fusión y ausencia de estructura secundaria utilizados para los oligonucleótidos diseñados, pero esta vez en regiones que no sean compartidas por ninguno de los genes de cada alineamiento. Al utilizar estas secuencias como sondas se aprovecharía de forma más eficiente el espacio disponible en un microarreglo, y al hallar puntos hibridados con una muestra podría decirse que la secuencia detectada tiene

mayor similitud con el gen de una especie determinada incluida en el microarreglo. La utilización de sondas diseñadas a partir de secuencias conservadas en una variedad de especies constituye un ensayo exploratorio aunque no excluyente, de forma que si una de estas sondas hibridiza con el DNA o el cDNA de la cepa investigada se tiene un indicio de la presencia del gen para el cual ha sido diseñada la sonda. Si no se presenta hibridización con las sondas diseñadas para una familia enzimática, no podría asegurarse que la cepa no posea genes que codifiquen para enzimas de dicha familia.

La clasificación de las cepas nativas a nivel de especie ha presentado dificultades, y sus caracteres taxonómicos únicos podrían conducir a su clasificación como una nueva especie dentro del género. Sin embargo, aún si se llegara a concluir que son cepas de la especie *C. butyricum*, el genoma de esta última no ha sido completamente secuenciado ni existen por el momento proyectos genoma que la cobijen. Por este motivo se consideró pertinente buscar secuencias de genes que sean característicos del género *Clostridium*, pero que también se encuentren en otros géneros relacionados.

Como se presentó en la metodología de construcción de la base de datos MULTI-CLOST, se incluyeron en ella registros especies bacterianas y fúngicas relacionadas con *Clostridium*, bien fuera por su clasificación taxonómica o por exhibir las mismas rutas metabólicas. Bajo el primer criterio se incluyeron especies de los géneros *Bacillus*, *Oceanobacillus*, *Geobacillus* y *Lactobacillus*, integrantes todos ellos del *Filum Firmicutes* (bacterias Gram (+)), así como otras bacterias del orden *Clostridiales* que exhiben las rutas metabólicas pertinentes, entre las que se cuentan varias especies de los géneros *Thermoanaerobacterium*, *Acetivibrio*, *Butyrivibrio*, *Ruminococcus* y *Caldicellulosiruptor* (Lynd *et al.*, 2002; Coughlan y Mayer, 1992; Rabinovich *et al.*, 2002; Tabla 1). Otros microorganismos alejados filogenéticamente del género *Clostridium* fueron incluidos por ser celulolíticos reconocidos, como es el caso de *Cellulomonas fimi*, *Hypocrea jecorina* y varias especies de los géneros bacterianos *Bacteroides* y *Fibrobacter* y de los hongos *Orpinomyces* y *Piromyces* (Coughlan y Mayer, 1992; Lynd *et al.*, 2002; Rabinovich *et al.*, 2002). Por su parte, las enterobacterias patógenas *Klebsiella pneumoniae* y *Citrobacter freundii* fueron incluidas por ser representativas de la producción de 1,3-PD a partir de glicerol (Sun *et al.*, 2003). Por último fue necesario incluir a las levaduras *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* y *Zygosaccharomyces rouxii*, considerando que son los únicos microorganismos que pueden convertir el glicerol-P proveniente de la glucólisis en glicerol (Biebl *et al.*, 1999).

Como se mostró en los resultados, las subcategorías de MULTI-CLOST con mayor número de enzimas condujeron a la obtención de una mayor cantidad de alineamientos y secuencias conservadas. Se escogió ClustalW 1.83 como *software* de alineamiento con los parámetros establecidos por defecto por ser lo suficientemente bajos para compensar la variación considerable entre los genes de cada familia enzimática a través de diferentes especies (Thompson *et al.*, 1994). La combinación de parámetros de alineamiento 65 - [15- 6,66] (PAP mínimo - [apertura- extensión de gaps]) en ClustalW fue aplicable a los grupos de genes que guardan una similitud considerable a lo largo de toda su secuencia y con ella se obtuvieron 16 alineamientos, la

mayor parte correspondientes a la categoría 'Solventogénesis'. El PAP promedio en estos alineamientos va de 66 a 94, y se conformaron por grupos de dos a cinco genes. Por su parte, la combinación 50 - [15- 6,66] fue aplicable a grupos de genes con PAP bajos pero que conservaban similitud significativa a través de toda la secuencia. Con esta combinación de parámetros se construyeron 30 alineamientos, siendo la combinación con mayor aplicabilidad para los genes analizados en este trabajo. Con ella se obtuvieron los mayores alineamientos, que contienen hasta siete genes cada uno, aunque los valores promedio de PAP descienden a un rango de 51 a 71,1. La combinación 50 - [11-1] constituyó la mejor opción en grupos cuyos genes comparten solo ciertos dominios funcionales, como es el caso de la mayoría de las celulasas multidominio (Bayer *et al.*, 1998). Con esta combinación se construyeron 11 alineamientos con PAP promedio que van de 55,9 a 77. A partir de las regiones conservadas en dichos alineamientos se diseñaron 94 sondas de oligonucleótidos. Uno de los principales sustentos teóricos para la búsqueda de secuencias conservadas en los genes estudiados es que en las especies correspondientes se han demostrado eventos de transferencia horizontal, siendo este de hecho el origen de varios operones en el genoma de *C. acetobutylicum* (Nölling *et al.*, 2001). En este estudio se dedujo la transferencia horizontal de la Endoglucanasa V de *Erwinia chrysantemi* a *Cellulomonas uda*, en donde se identifica como Endoglucanasa I, a partir de sus semejanzas en cuanto a secuencia nucleotídica, frecuencia de uso de codones y porcentaje de G+C. Sin embargo, los genes provenientes de especies fúngicas y actinomicetos almacenados en MULTI-CLOST formaron casi invariablemente alineamientos separados de aquellos en los cuales se concentraban genes de las clases *Clostridia* y *Bacilli*, aun con el uso de parámetros de alineamiento poco restrictivos (datos no mostrados). Este resultado contrasta con la hipótesis adelantada por García-Vallve *et al.* (2000), en la que se propone que los genes de celulasas en hongos del rúmen fueron obtenidos mediante transferencia horizontal a partir de bacterias del rúmen (hipótesis reforzada por la ausencia de intrones en dichos genes). La base de datos MULTI-CLOST y las sondas de oligonucleótidos diseñadas son herramientas muy útiles para el mejoramiento futuro de las cepas nativas de *Clostridium* sp. por ingeniería metabólica, puesto que permiten analizar y comprender las rutas metabólicas implicadas en el consumo de sustratos de bajo costo y en la producción de metabolitos de alto valor comercial en diferentes especies. Con MULTI-CLOST se obtiene una visión amplia de las especies que representan cada ruta metabólica, así como de las variaciones enzimáticas que existen entre ellas, mientras que con las sondas de oligonucleótidos se puede analizar el perfil de expresión de los genes implicados en cada ruta. La estrategia propuesta para el diseño de oligonucleótidos guarda semejanza con la estrategia propuesta en la patente de Graham y Wollweber (1990) para la detección de genes de α -Amilasa en diferentes especies del género *Bacillus*, así como con la estrategia diseñada por Matveeva *et al.* (2004) para el análisis de variantes del virus VIH-1 por ensayos de hibridación. Sin embargo, la estrategia para el análisis de rutas metabólicas en las cepas colombianas de *Clostridium* sp. muestra ventajas con respecto a la especificidad de detección y a la cantidad de sondas obtenidas. En la estrategia patentada por Graham y Wollweber (1990), por ejemplo, se recurre al uso de nucleótidos ambiguos o bien se escogen nucleótidos arbitrarios en las posiciones sin consenso de los alineamientos múltiples,

arriesgando de esta forma la especificidad y capacidad de detección de las sondas resultantes. Por otra parte la estrategia de Matveeva *et al.* (2004) es muy similar a la de este artículo en cuanto a la selección de sondas a partir de los alineamientos múltiples por parámetros de longitud, valores termodinámicos y formación de estructuras secundarias o dímeros, pero su estructura automatizada y exhaustiva resulta poco práctica para el diseño de un número limitado de sondas que cubran un amplio espectro de genes. La estrategia utilizada en el presente estudio resultó eficaz para el diseño de sondas dirigidas a regiones consenso de cada gen objetivo, con las cuales se espera analizar en el futuro las rutas metabólicas de interés en las cepas colombianas de *Clostridium* sp. y así mejorar eventualmente su rendimiento industrial.

AGRADECIMIENTOS

Este trabajo fue desarrollado dentro del marco del proyecto “Producción de 1,3 Propanodiol a partir de glicerol generado del proceso productivo de Biodiesel, empleando cepas nativas de *Clostridium* spp.: estudio del operón condiciones de producción por fermentación y su viabilidad económica”, código 1101-12-17848 financiado por el Instituto Colombiano Francisco José de Caldas COLCIENCIAS y la Universidad Nacional de Colombia.

BIBLIOGRAFÍA

- AIYAR A. The Use of CLUSTAL W and CLUSTAL X for Multiple Sequence Alignment. En: Misener S, Krawetz S, editors. Bioinformatics: Methods and protocols. New Jersey: Humana Press Inc.; 2000. p. 221-239.
- ALSAKER K, PAPOUTSAKIS E. Transcriptional Program of Early Sporulation and Stationary-Phase Events in *Clostridium acetobutylicum*. J Bacteriol. 2005;187(20):7103-7118.
- ALTSCHUL S, MADDEN T, SCHÄFFER A, ZHANG J, ZHANG Z, MILLER W, *et al.* Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. Nucleic Acids Res. 1997;25(17):3389-3402.
- ARÉVALO C. Contribución a la caracterización molecular de cepas nativas de *Clostridium* spp. con potencial biotecnológico [tesis de maestría]. Bogotá D.C.: Posgrado Interfacultades en Microbiología, Universidad Nacional de Colombia; 2001.
- BAIROCH A, APWEILER R, WU C, BARKER W, BOECKMANN B, FERRO S, *et al.* The Universal Protein Resource (UniProt). Nucleic Acids Res. 2005;33:D154-D159.
- BAYER E, SHIMON L, SHOHAM Y, LAMED R. Cellulosomes-Structure and Ultrastructure. J Struct Biol. 1998;124(2-3):221-234.
- BENSON D, KARSCH-MIZRACHI I, LIPMAN D, OSTELL J, WHEELER D. GenBank. Nucleic Acids Res. 2007;35(Database issue):D21-D25.
- BIEBL H, MENZEL K, ZENG A, DECKWER W. Microbial Production of 1,3-Propanediol. Appl Microbiol Biotechnol. 1999;52(3):289-297.
- BUNNEY W, BUNNEY B, VAWTER M, TOMITA H, LI J, EVANS S, *et al.* Microarray Technology: A Review of New Strategies to Discover Candidate Vulnerability Genes in Psychiatric Disorders. Am J Psychiatry. 2003;160(4):657-666.

CÁRDENAS D, PULIDO C, ARAGÓN O, ARISTIZÁBAL F, SUÁREZ Z, MONTOYA D. Evaluación de la producción de 1,3-propanodiol por cepas nativas de *Clostridium* sp. mediante fermentación a partir de glicerol USP y glicerol industrial subproducto de la producción de Biodiesel. *Rev Col Cienc Quím Farm.* 2006;35(1):120-137.

COUGHLAN M, MAYER F. The Cellulose-Decomposing Bacteria and their Enzyme Systems. En: Balows A, editor. *The prokaryotes: A Handbook on the Biology of Bacteria.* 2 ed. New York: Springer-Verlag; 1992. p. 460-502.

CRONIN M, PHO M, DUTTA D, FRUEH F, SCHWARCZ L, BRENNAN T. Utilization of New Technologies in Drug Trials and Discovery. *Drug Metab Dispos.* 2001;29(4):586-590.

GARCÍA-VALLVE S, ROMEU A, PALAU J. Horizontal Gene Transfer of Glycosyl Hydrolases of the Rumen Fungi. *Mol Biol Evol.* 2000;17(3):352-361.

GRAHAM B, WOLLWEBER K. Oligonucleotide Probes for Detection of Alpha-Amylase Genes. Washington D.C.: U.S. Patent and Trademark Office; 1990. U.S. Patent 4917999.

GUCKENBERGER M, KURZ S, AEPINUS C, THEISS S, HALLER S, LEIMBACH T, et al. Analysis of the Heat Shock Response of *Neisseria meningitidis* with cDNA- and Oligonucleotide-Based DNA Microarrays. *J Bacteriol.* 2002;184(9):2546-2551.

HARRIS L, WELKER N, PAPOUTSAKIS E. Northern, Morphological, and Fermentation Analysis of *spo0A* Inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J Bacteriol.* 2002;184(13):3586-3597.

HEGDE P, QI R, ABERNATHY K, GAY C, DHARAP S, GASPARD R, et al. *A Concise Guide to cDNA Microarray Analysis.* Biotechniques. 2000;29(3):548-562.

JAIMES C, ARISTIZÁBAL F, BERNAL M, SUÁREZ Z, MONTOYA D. AFLP Fingerprinting of Colombian *Clostridium* spp. Strains, Multivariate Data Analysis and Its Taxonomical Implications. *J Microbiol Methods.* 2006;67(1):64-69.

JONES D, WOODS D. Acetone-Butanol Fermentation Revisited. *Microbiol Rev.* 1986;50(4):484-524.

LEE I, DOMBKOWSKI A, ATHEY B. Guidelines for Incorporating Non-Perfectly Matched Oligonucleotides Into Target-Specific Hybridization Probes for a DNA Microarray. *Nucleic Acids Res.* 2004;32(2):681-690.

LUCAS R, ROBLES A, GÁLVEZ A, GARCÍA T, PÉREZ R, ÁLVAREZ G. Biodegradación de la celulosa y la lignina. Jaén: Junta Andalucía Consejería EDU; 2001. p. 14-62.

LYND L, WEIMER P, VAN ZYL W, PRETORIUS I. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol Mol Biol Rev.* 2002;66(3):506-527.

MATVEEVA O, FOLEY B, NEMTSOV V, GESTELAND R, MATSUFUJI S, ATKINS J, et al. Identification of Regions in Multiple Sequence Alignments Thermodynamically Suitable for Targeting by Consensus Oligonucleotides: Application to HIV Genome. *BMC Bioinformatics* [serial online] 2004 [citado 5 Sep 2007]; 5(44): [15 pantallas]. Disponible en: URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=419695>.

MONTOYA D, SPITIA S, SILVA E, SCHWARZ W. Isolation of Mesophilic Solvent-Producing Clostridia from Colombian Sources: Physiological Characterization, Solvent Production and Polysaccharide Hydrolysis. *J Biotechnol.* 2000;79(2):117-126.

MONTOYA D, ARÉVALO C, GONZALES S, ARISTIZÁBAL F, SCHWARZ W. New Solvent-Producing *Clostridium* sp. Strains, Hydrolyzing a Wide Range of

Polysaccharides, are Closely Related to *Clostridium butyricum*. J Ind Microbiol Biotechnol. 2001;27(5):329-335.

MONTOYA J, SUÁREZ Z, MONTOYA D, ARISTIZÁBAL F. Análisis bioinformático y predicción de genes en secuencias genómicas de *Clostridium* sp. IBUN22A. Rev Col Biotecnol. 2006;8(1):57-64.

NAKAMURA C, WHITED G. Metabolic Engineering for the Microbial Production of 1,3-Propanediol. Curr Opin Biotechnol. 2003;14(5):454-459.

NICHOLAS K, NICHOLAS H. GeneDoc: A Tool for Editing and Annotating Multiple Sequence Alignments [programa de ordenador *online*]. Versión 2.6.002. Pittsburg (PA): Pittsburgh Supercomputing Center; 2001. Disponible en: URL: <http://www.psc.edu/biomed/genedoc/>.

NÖLLING J, BRETON G, OMELENKO M, MAKAROVA K, ZENG Q, GIBSON R, *et al.* Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*. J Bacteriol. 2001;183(16):4823-4838.

PACHAURI N, HE B. Value-Added Utilization of Crude Glycerol from Biodiesel Production: A Survey of Current Research Activities. Portland: 2006 ASABE Annual International Meeting; 2006. Paper Number: 066223.

PAREDES C, SENGER R, SPATH I, BORDEN J, SILLERS R, PAPOUTSAKIS E. A General Framework for Designing and Validating Oligomer-Based DNA Microarrays and its Application to *Clostridium acetobutylicum*. Appl Environ Microbiol. 2007;73(14):4631-4638.

PELÁEZ T, ALCALÁ L, ALONSO R, RODRÍGUEZ-CRÉIXEMS M, GARCÍA-LECHUZ J, BOUZA E. Reassessment of *Clostridium difficile* Susceptibility to Metronidazole and Vancomycin. Antimicrob Agents Chemother. 2002;46(6):1647-1650.

RABINOVICH M, MELNICK M, BOLOBOVA A. The Structure and Mechanism of Action of Cellulolytic Enzymes. Biochemistry Moscow. 2002;67(8):850-871.

RUTHERFORD K, PARKHILL J, CROOK J, HORSNELL T, RICE P, RAJANDREAM M, *et al.* Artemis: Sequence Visualisation and Annotation. Bioinformatics. 2000;16(10):944-945.

RYBICKI E. A Manual of Online Molecular Biology Techniques. Cape Town: University of Cape Town; 2005. Disponible en: URL: <http://www.mcb.uct.ac.za/manual/MolBiolManual.htm>.

SANTALUCIA J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. Proc Natl Acad Sci USA. 1998;95:1460-1465.

SCHUCHHARDT J, BEULE D, MALIK A, WOLSKI E, EICKHOFF H, LEHRACH H, *et al.* Normalization Strategies for cDNA Microarrays. Nucleic Acids Res. 2000;28(10):e47i-e47v.

SCHWARZ, W. The Cellulosome and Cellulose Degradation by Anaerobic Bacteria. Appl Microbiol Biotechnol. 2001;56(5-6):634-649.

STOUGHTON R. Applications of DNA Microarrays in Biology. Annu Rev Biochem. 2005;74:53-82.

SUÁREZ Z. Contribución a la taxonomía de 13 cepas nativas de *Clostridium* sp. mediante análisis multivariado de técnicas de caracterización genotípicas y fenotípicas [tesis de maestría]. Bogotá D.C.: Posgrado Interfacultades en Microbiología, Universidad Nacional de Colombia.; 2004.



SUN J, VAN DER HEUVEL J, SOUCAILLE P, QU Y, ZENG A, *et al.* Comparative Genomic Analysis of Dha Regulon and Related Genes for Anaerobic Glycerol Metabolism in Bacteria. *Biotechnol Prog.* 2003;19(2):263-272.

THOMPSON J, HIGGINS D, GIBSON T. Clustal W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* 1994;22(22):4673-4680.

TOLSTRUP N, NIELSEN P, KOLBERG J, FRANKEL A, VISSING H, KAUPPINEN S. OligoDesign: Optimal Design of LNA (Locked Nucleic Acid) Oligonucleotide Capture Probes for Gene Expression Profiling. *Nucleic Acids Res.* 2003;31(13):3758-3762.

TOMAS C, ALSAKER K, BONARIUS H, HENDRIKSEN W, YANG H, BEAMISH J, *et al.* DNA Array-Based Transcriptional Analysis of Asporogenous, Nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *J Bacteriol.* 2003;185(15):4539-4547.

TUMMALA S, JUNNE S, PAPOUTSAKIS E. Antisense RNA Downregulation of Coenzyme A Transferase Combined with Alcohol-Aldehyde Dehydrogenase Overexpression Leads to Predominantly Alcohologenic *Clostridium acetobutylicum* Fermentations. *J Bacteriol.* 2003;185(12):3644-3653.

VELCULESCU V, ZHANG L, VOGELSTEIN B, KINZLER K. Serial Analysis of Gene Expression. *Science.* 1995;270:484-487.

VARGAS C, MONTOYA D, ARISTIZÁBAL F. Clonación y expresión en *Escherichia coli* de genes de celulasas de *Clostridium* IBUN 22A. *Rev Col Biotechnol.* 2002;4(1):29-35.

ZVERLOV V, BEREZINA O, VELIKODVORSKAYA G, SCHWARZ W, *et al.* Bacterial Acetone and Butanol Production by Industrial Fermentation in the Soviet Union: Use of Hydrolyzed Agricultural Waste for Biorefinery. *Appl Microbiol Biotechnol.* 2006;71(5):587-597.