

03081

6
2ej

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Instituto de Biotecnología

**Análisis del sitio activo de enzimas mediante
mutagénesis combinatoria**

Tesis

**que para obtener el grado de
Doctor en Investigación Biomédica Básica**

presenta

el M. en IBB. Joel Osuna Quintero

**TESIS CON
FALLA DE ORIGEN**

Cuernavaca, Morelos

Abril 1992

**TESIS CON
FALLA DE ORIGEN**



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Índice

Abstract	2
Resumen	3
Introducción	4
Proteínas modelo	8
EcoRI	8
β -lactamasa	10
Mutagénesis combinatoria	14
Resultados	15
Mutagénesis de la β -lactamasa	15
Clonación y electroporación del banco de genes mutantes.	17
Discusión	19
EcoRI	19
β -lactamasa	20
1. transformación eficiente de la mezcla de genes mutantes.	21
2. contaminación con el gene silvestre.	21
Conclusiones	23
Artículo 1.º	24
Artículo 2.º	31
Referencias	35

Abstract

In the study of the structure-function relationship of proteins, as well as for the generation of proteins with novel properties, methods for the site-directed mutagenesis of the corresponding genes have become an indispensable tool. Recent literature shows a number of experimental approaches for oligonucleotide-directed, site-specific mutagenesis, including those employing PCR. The genetic approaches to engineering a new function or to identify the specificity or catalytic determinants in an enzyme are the more powerful alternatives to the design of novel properties in proteins. Once a handful of key residues have been identified for an enzyme, a library of mutant enzymes with all combinations of random substitutions at these positions can be generated. The development of sensitive and easy screenings or selections schemes for isolating interesting enzymes, from the uninteresting or non-functional mutants, is the major problem to this approach. In this thesis we describe methods to combinatorially mutagenize several residues that are far away from each other in the sequence but close together in the active site of the protein. The model proteins used were the *EcoRI* restriction endonuclease and the RTEM- β -lactamase.

With *EcoRI*, a library of mutant genes was generated using trinucleotide blocks and a combination of recombinant DNA procedures, including primer extension and the PCR. Codons corresponding to three amino acids previously implicated in the specific recognition of the DNA substrate, were combinatorially mutated so as to generate a library that potentially contains all 20^3 possible single, double and triple amino acid replacements, in a balanced distribution. Inspection of the phenotypes of *Escherichia coli* colonies bearing the mutant genes showed that several of them retained activities that were deleterious to the cells but were still protected by the *EcoRI* methylase. These included new enzymes variants, including non-conservative single and double replacements.

In the case of β -lactamase, we combinatorially mutated six residues close to the active-site serine of the enzyme. With this model protein, the mutagenesis strategy was based on PCR. The scheme is suitable for generating simultaneous, noncontiguous mutations on the same molecule with high efficiency. At the moment, we have a library of variant β -lactamase genes. Our goal is to construct a new residue constellation that assists to the active-site serine in the serin- β -lactamase and to find out the residue constellation needed to assist the active-site cysteine of the thiol- β -lactamase.

Resumen

Los métodos de mutagénesis sitio-dirigida se han convertido en una herramienta indispensable para el análisis de la estructura-función de una proteína, así como para la generación de proteínas con funciones ó propiedades novedosas. Recientemente han aparecido un gran número de enfoques de mutagénesis sitio-dirigida basada en oligonucleótidos sintéticos, incluyendo métodos que utilizan PCR. La alternativa genética para el desarrollo de una nueva función ó para la identificación de los determinantes de especificidad ó de catálisis de una enzima, es sin duda alguna un enfoque poderoso para el diseño de propiedades novedosas en proteínas. Una vez que se han identificado los residuos importantes para la función de una enzima, puede producirse un banco de enzimas mutantes con todas las posibles combinaciones de sustituciones azarosas en esas posiciones. El principal problema de este enfoque consiste en la necesidad de ensayos de búsqueda, simples y eficientes, de aquellas variantes interesantes. En el presente trabajo de tesis, describimos métodos para mutagenizar en forma combinatoria varios residuos lejanos en la secuencia de aminoácidos de una proteína, pero cercanos en el sitio activo. Como proteínas modelo utilizamos a la endonucleasa de restricción *EcoRI* y a la RTEM- β -lactamasa.

Con *EcoRI*, se construyó un banco de genes mutantes utilizando trinucleótidos sintéticos y una combinación de metodologías de DNA recombinante, incluyendo extensión de primer y PCR. Se mutagenizaron en forma combinatoria a los codones correspondientes a los tres residuos previamente implicados en el reconocimiento específico del sustrato, generando, de esta manera, un banco que potencialmente contiene los 20^3 posibles reemplazos sencillos, dobles y triples en una distribución balanceada. La inspección de los fenotipos de colonias de *Escherichia coli* portadoras de los genes mutantes ha mostrado variantes que retienen actividad deletérea para la célula pero sensible a la metilasa *EcoRI*. Esas enzimas mutantes incluyen nuevas variantes con reemplazos no conservativos sencillos o dobles.

En el caso de la β -lactamasa, hemos mutado en forma combinatoria 6 residuos cercanos a la serina del sitio activo de la enzima. Con esta proteína modelo, se utilizó una estrategia de mutagénesis basada en PCR. El esquema de mutagénesis es apropiado para la generación eficiente de mutaciones simultáneas, no contiguas, en la misma molécula. Hasta el momento, se cuenta con un banco de genes mutantes de la β -lactamasa. Nuestra meta es construir una nueva constelación de residuos que asista a la serina del sitio activo en la serin- β -lactamasa y además, encontrar la constelación de residuos apropiados para asistir a la cisteína del sitio activo de la tiol- β -lactamasa.

Introducción

El rápido desarrollo en técnicas de Biología Molecular ha impulsado notablemente la ingeniería de proteínas y ha permitido que esta disciplina se convierta en un medio único para examinar el sitio activo de enzimas, de tal forma que se pueda identificar y cuantificar el papel de residuos específicos en la unión, especificidad y catálisis de sustratos.

Dada la complejidad de las interacciones que mantienen la estructura del sitio activo, es difícil predecir el o los cambios necesarios para producir una enzima con nuevas capacidades catalíticas. El desarrollo de nuevos procedimientos computacionales capaces de calcular y simular los efectos de mutagénesis dirigida en la estabilidad y actividad de enzimas y complejos enzima-sustrato, sugieren la posibilidad de predecir las consecuencias funcionales de sustituciones de residuos que forman el sitio activo de la enzima(1,2). A pesar de lo prometedoras que son estas estrategias racionales de modificación de enzimas, hasta el momento han tenido un éxito limitado a sistemas enzima o enzima-sustrato excepcionalmente caracterizados tanto estructural como bioquímicamente(1,2).

En este trabajo proponemos una alternativa al enfoque anterior de diseño de enzimas. Nuestra alternativa consiste en la modificación a saturación y en una forma azarosa de combinaciones, de todos los residuos potencialmente involucrados en las capacidades de unión o de catálisis de la enzima. La existencia de fenómenos de aditividad de diferentes reemplazos en una proteína, así como la posibilidad de interacciones

complejas entre residuos aminoácidos vecinos en la estructura tridimensional de una proteína (3) sirve como justificación para este enfoque combinatorio de mutagénesis. En el caso de mutagénesis aditiva, es claro que la combinación de mutantes sencillas, cada una de las cuales aporta una pequeña mejoría a la función deseada, es una de las herramientas más empleada en ingeniería de proteínas(4-9). La mutagénesis no-aditiva ocurre con residuos catalíticos interdependientes, la modificación de uno de ellos puede dar como resultado que los demás obstaculicen la función. También en estos casos es recomendable la modificación simultánea de este subgrupo de residuos para lograr el cambio a la función deseada(10). El éxito de la mutagénesis combinatoria sólo depende del desarrollo de técnicas de búsqueda o selección, rápidas y eficientes, de la función deseada y de la eficiencia de la mutagénesis(11). La posibilidad de expresar proteínas en la superficie de fagos filamentosos (12-14) sin duda alguna será de gran utilidad para el tipo de estrategias selectivas. Mediante purificación por afinidad se puede seleccionar la proteína expresada con base en la especificidad de unión con el ligando. Si somos capaces de producir un banco de proteínas mutantes expresadas en el fago filamentosos es posible analizar millones de estas mutantes y seleccionar solo aquellas con mayor especificidad por el ligando.

Existe un gran número de estrategias diferentes que permiten modificar o construir genes que codifiquen para la secuencia de aminoácidos que deseamos. De hecho, gracias a la comercialización de algunos de estos métodos, en la actualidad es relativamente fácil adquirir esta habilidad.

La mayoría de los métodos dirigidos a la modificación de regiones específicas de alguna proteína utilizan oligonucleótidos sintéticos que codifican para el o los cambios planeados. El enfoque azaroso de mutagénesis dirigida depende de oligonucleótidos sintéticos "degenerados" en las bases que codifican para el o los aminoácidos que deseamos cambiar. Esto se logra gracias a la versatilidad que la automatización permite durante la síntesis del oligonucleótido. Más adelante discutiremos con mas detalle como se lleva a cabo la producción de estos oligonucleótidos, por el momento, nos limitaremos a decir que nuestra estrategia de mutagénesis depende de varios de estos oligonucleótidos "degenerados" complementarios a diferentes regiones del gene que nos interesa modificar.

A pesar de la gran cantidad de métodos diferentes de mutagénesis dirigida, solo existe un número muy limitado de estrategias que permiten la modificación simultánea de residuos lejanos en la secuencia primaria de la proteína pero cercanos en la estructura tridimensional. La gran mayoría de los ejemplos de mutagénesis combinatoria consisten en la modificación simultánea de residuos cercanos en secuencia, esto permite un diseño de mutagénesis que utiliza uno o dos oligonucleótidos que son portadores de los cambios necesarios(15-18). La mayoría de los ejemplos de mutagénesis combinatoria de residuos lejanos en secuencia no emplean mutagénesis dirigida en forma azarosa, sino que dependen de dos o más ciclos de mutagénesis sencilla de tal forma que pueden producirse mutantes dobles o múltiples de los residuos blanco(4). Existen estrategias interesantes de mutagénesis al azar dirigidas a residuos lejanos que utilizan la síntesis completa de la región blanco del gene(19). Los métodos de mutagénesis más eficientes

(teóricamente 100% de eficiencia) emplean la reacción en cadena de la polimerasa (PCR). Recientemente, Nelson y Long (20) reportaron una versión muy novedosa de mutagénesis dirigida por PCR. El empleo de un oligonucleótido adicional ("tapón") el cual no es complementario a la secuencia templado original permite extender este esquema a la combinación de cada producto mutante de PCR con mutaciones codificadas en el siguiente oligonucleótido, con la ventaja adicional de poder ser selectivo para la amplificación exclusiva de secuencias mutantes, evitándose de esta manera la producción del gene silvestre. Más adelante explicaremos con más detalle las diferentes etapas para la producción de mutaciones múltiples combinatorias por PCR.

Proteínas modelo

EcoRI

La endonucleasa *EcoRI* es la enzima de restricción más caracterizada debido a su importancia intrínseca y a las consecuencias que su estudio pudiera tener sobre generalidades de la interacción proteína-DNA.

El primer reporte de la estructura del complejo enzima-DNA (21) propuso tres residuos aminoácidos como los responsables de la especificidad de la enzima por el sitio 5'GAATTC 3' del DNA.

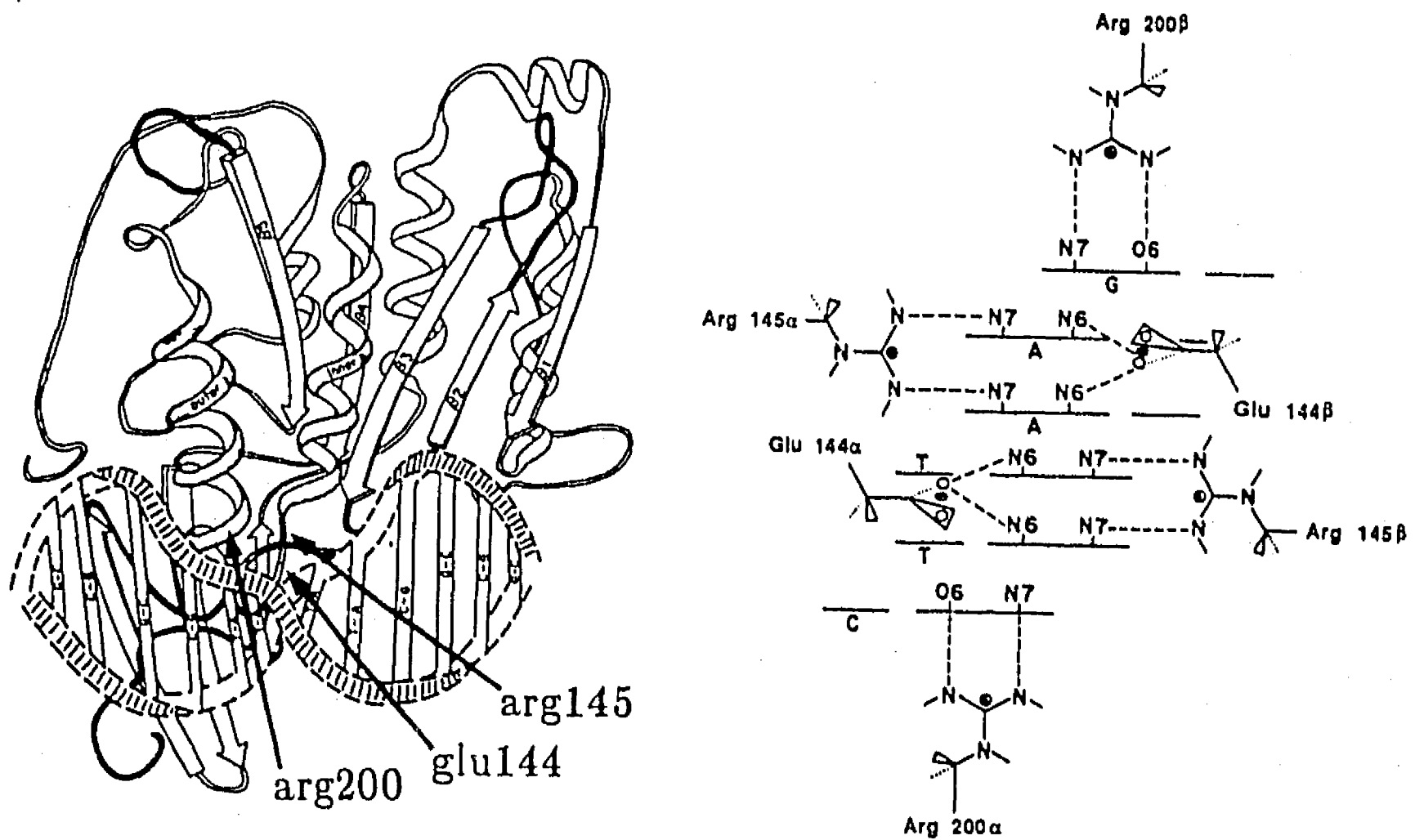


Fig. 1 Estructura propuesta para el complejo *EcoRI*-DNA a 3 Å de resolución (21). A la izquierda, se muestra un monómero de la *EcoRI*. Cada monómero presenta dos α -hélices (externa e interna) que proyectan hacia el surco mayor del DNA. Las flechas indican la posición de los tres aminoácidos propuestos para la interacción con los nucleótidos del sitio de reconocimiento. A la derecha, un detalle de los puentes de hidrógeno propuestos entre la proteína y el DNA.

En ese momento, se llegó a la conclusión que la especificidad de secuencia estaba mediada por 12 puentes de hidrógeno formados directamente entre 3 aminoácidos y las bases purínicas de la secuencia canónica. Los residuos involucrados en las interacciones con el DNA fueron los siguientes: Glu144, Arg145 y Arg200 (fig.1). Una revisión reciente de la estructura del complejo (22) ha revelado regiones adicionales de la proteína que contactan las bases pirimidínicas de la secuencia canónica. En este nuevo modelo, ya refinado, la especificidad de secuencia parece estar determinada por 16 puentes de hidrógeno entre aminoácidos de la proteína y las bases del DNA (fig.2), además de una red de contactos tipo Van der Waals que involucra a todas las pirimidinas.(23)

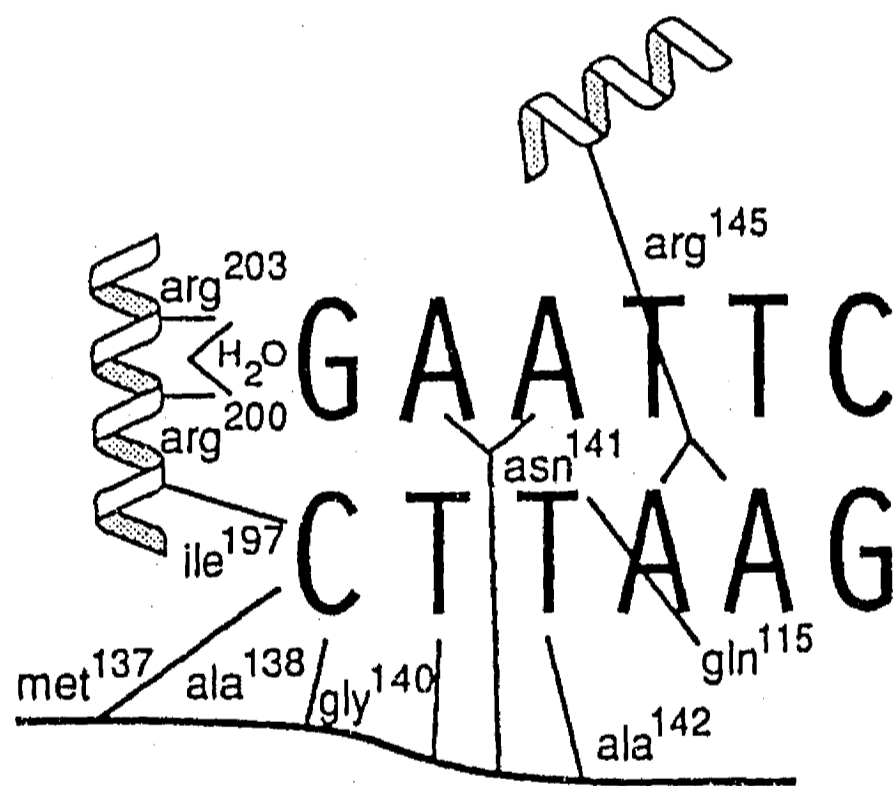


Fig. 2 Modelo del reconocimiento del sustrato por la *EcoRI* según la revisión de la estructura del complejo enzima-DNA (22). Además de las hélices externa (resíduos 197, 200 y 203) e interna (resíduo 145), cada monómero presenta un segmento de cadena polipeptídica en forma extendida, el cual atraviesa el surco mayor del DNA y contacta algunas bases del sitio de reconocimiento.

La actividad letal, *in vivo*, de la enzima silvestre y la existencia de una metilasa *EcoRI* que protege de manera efectiva al DNA contra la actividad de la endonucleasa, nos ha permitido diseñar un ensayo *in vivo* lo suficientemente sensible para detectar enzimas mutantes con actividad normal o alterada contra el sitio canónico y enzimas mutantes con diferente especificidad de reconocimiento.

En el inicio de este trabajo nos planteamos los siguientes objetivos: analizar el contenido informacional de cada uno de los tres residuos que aparentemente eran responsables de la especificidad y producir una combinación de mutaciones a saturación en esas tres posiciones con el fin de poder encontrar variantes de la enzima con actividad catalítica contra secuencias de DNA diferentes a la canónica.

β -lactamasa

Pocas enzimas tienen tantos atributos para el ingeniero de proteínas como las β -lactamasas. Existen por lo menos 4 grupos diferentes de β -lactamasas: A, B, C y D, clasificados de acuerdo a su identidad de secuencia, a su especificidad de sustrato y a su comportamiento cinético. Exceptuando a la clase B, todas las demás β -lactamasas son hidrolasas que utilizan una serina como residuo catalítico. Aunque pudiera pensarse en una similitud a las serin-proteasas, las β -lactamasas carecen del residuo catalítico histidina que es común a todas las serin-proteasas, son incapaces de hidrolizar péptidos y su mecanismo de catálisis no parece ser idéntico al de las serin-proteasas. Las β -lactamasas hidrolizan e inactivan antibióticos β -lactámicos tales como penicilinas y cefalosporinas. Entre las β -lactamasas con serina-catalítica, las de clase A son las más

caracterizadas y estudiadas: se conoce la estructura tridimensional de al menos 4 enzimas de esta clase (24-27), se tiene la secuencia de aminoácidos de por lo menos 20 enzimas de clase A (28) y se han realizado un gran número de experimentos relacionados con la identificación de residuos involucrados tanto en catálisis como en especificidad por el sustrato (29-31). El análisis de la estructura ha mostrado que existe una arquitectura extremadamente similar del sitio activo y de la topología en general (fig.3).

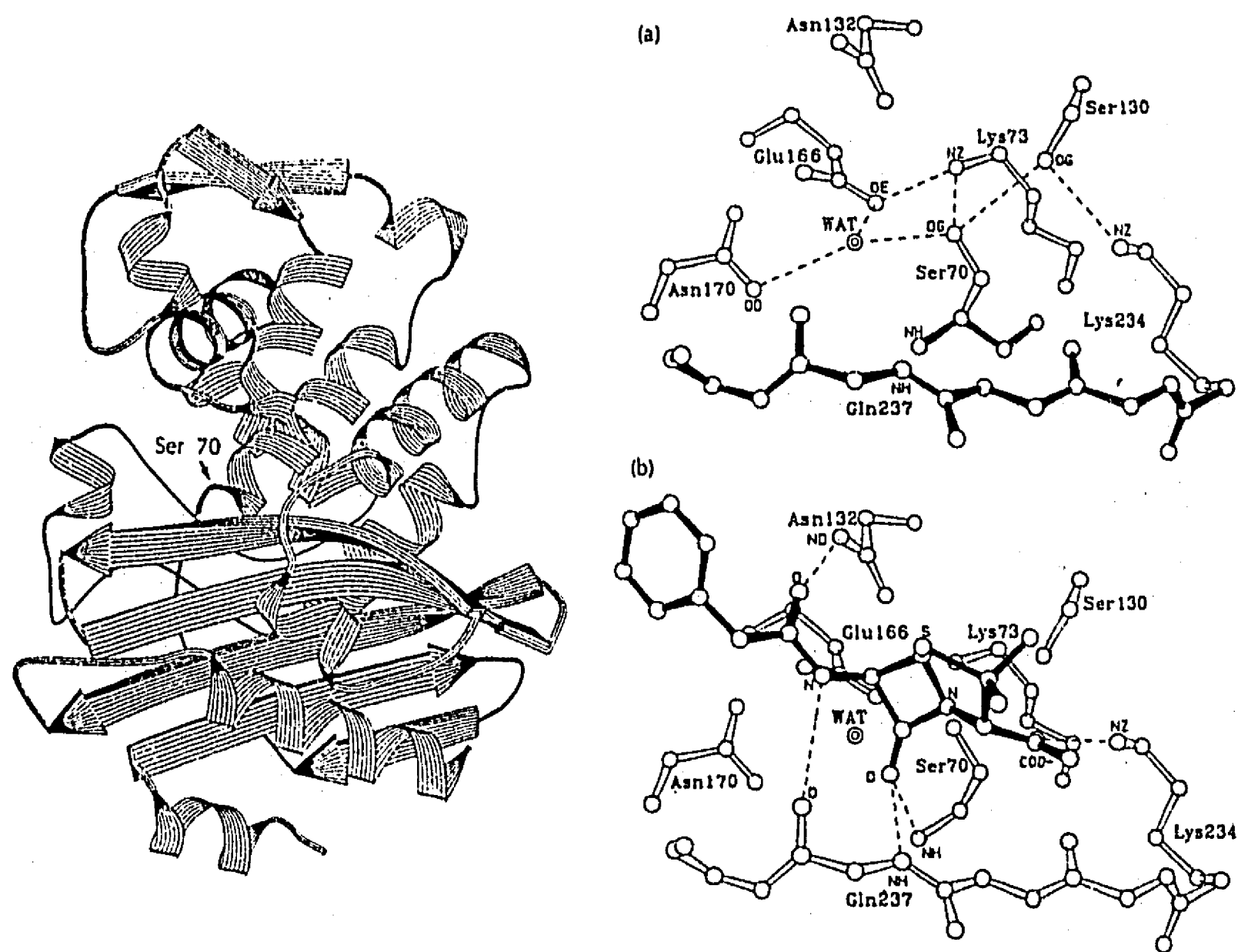


Fig. 3 A la izquierda, tipo de topología de las β -lactamasas de clase A. Se indica la posición de la serina del sitio activo. A la derecha, residuos clave del sitio activo de la β -lactamasa de *Staphylococcus aureus* PC1, sin sustrato (a) y con sustrato (b). Algunas Interacciones importantes se muestran con líneas discontinuas. WAT, es la molécula de agua necesaria para la hidrólisis.

Estos estudios han sido relevantes para la identificación de los posibles residuos que auxilian a la serina-catalítica en su función. Haciendo analogía con las serin-proteasas se ha sugerido (32) que existe un mecanismo similar para estabilizar la carga negativa del intermediario del estado de transición. Se ha postulado que el residuo lisina en la posición 73 pudiera estar involucrado en el mecanismo de transferencia del protón de la serina activa hacia el sustrato, aunque no se descarta la idea de que la serina 130 pudiera estar relacionada también con este mecanismo. Para el paso de desacilación se ha propuesto la mediación de una molécula de solvente ayudada por el residuo glutámico 166 (fig.3). Debido a la carencia de un análisis estructural de un complejo enzima-sustrato que permita identificar de manera exacta un mecanismo probable de catálisis, varios grupos de trabajo se han interesado en las estrategias de mutagénesis dirigida como una alternativa para conocer la importancia de los residuos del sitio activo y el probable mecanismo catalítico de la enzima(33-42). Uno de los primeros trabajos relacionados con este tipo de caracterización fué la construcción de la tior- β -lactamasa (43), donde un grupo tior de cisteína reemplaza al grupo hidroxilo de la serina como nucleófilo primario de la catálisis. La existencia de tior-proteinasas análogas a las serin-proteasas (44) ha sugerido la posibilidad evolutiva de las tior- β -lactamasas o bien la posibilidad de intentar su diseño. La aproximación más cercana a su construcción (43) no tomó en consideración al resto de la proteína, y en particular al resto de los residuos del sitio activo, los cuales, en principio, pudieran estar evolutivamente optimizados para una serin- β -lactamasa.

En nuestro laboratorio estamos interesados en participar en la identificación de los residuos importantes para la catálisis así como para la especificidad de sustrato de la β -lactamasa de clase A. Creemos que el enfoque combinatorio de mutagénesis ya mencionado, es ideal para la caracterización de los residuos indispensables para la función de la enzima y en particular para analizar la posibilidad de producir nuevas constelaciones funcionales de residuos alrededor de la serina-activa o de encontrar la constelación de residuos activos necesarios para auxiliar a la cisteína-activa de la tior- β -lactamasa.

Mutagénesis combinatoria

En el inicio de este trabajo, como se muestra en la figura 1 del artículo 1, decidimos implementar un enfoque de mutagénesis combinatoria sin PCR. Tomando como base los resultados de análisis teóricos que proponen al codón como blanco de mutagénesis (45), decidimos implementar una estrategia de mutagénesis a saturación empleando trinucleótidos sintéticos como unidades mutagénicas (fig.1 del artículo 1).

El empleo de PCR para mutagénesis nos ha suministrado una mayor versatilidad de diseño de estrategias, permitiendo además, las altas eficiencias logradas en la mutagénesis combinatoria reportada en el artículo 1. Una ventaja de este tipo de mutagénesis consiste en la independencia de sitios de restricción cercanos a los residuos blanco de la mutagénesis. El empleo de oligonucleótidos sintéticos no complementarios a la secuencia templado original (20), nos ha permitido combinar productos mutantes de PCR con oligos mutantes adicionales evitándose la producción de secuencias silvestres (fig.1 del artículo 2).

Resultados

Los resultados obtenidos en este trabajo relacionados con la proteína modelo *EcoRI* se encuentran descritos en el artículo 1. Parte de los resultados relacionados con la β -lactamasa, principalmente los que involucran estrategia de mutagénesis se encuentran en prensa y se describen en el artículo 2 (por razones técnicas, en este artículo se utilizó a la *EcoRI* como modelo para describir la estrategia de mutagénesis combinatoria utilizando PCR). El resto de esta sección se utilizará para describir algunos resultados obtenidos (todavía no publicados) con la proteína modelo β -lactamasa.

Mutagénesis de la β -lactamasa

Con el fin de mutagenizar las posiciones del sitio activo ya mencionadas se sintetizaron los siguientes oligonucleótidos:

posición 234: 5'-GGTTTATTGCTGATNNG/CTCTGGAGCCGGTG-3'

posición 166: 5'-GCCTTGATCGTTGGNNG/CCCGGAGCTGAATG-3'

posición 130-132: 5'-CTGCCATAACCATGNNG/CNNG/CNNG/CACTGCGGCCAAC-3'

posición 73(S70): 5'-CAATGATGAGCACTTTTNG/CGTTCTGCTATGTGGC-3'

posición 73(C70): 5'-CAATGATGTGCACTTTTNG/CGTTCTGCTATGTGGC-3'

oligo A: 5'-GGAATAGCCGATCGAGATCTTGAGTAACTTGGTC-3'

oligo B (tapón): 5'-GGAATAGCCGATCGAGATCT-3'

oligo C: 5'-GCAATTTAACTGTGATAAACTACCGC-3'

Durante la síntesis del oligonucleótido, el o los codones blanco de la mutagénesis se reemplazaron por **NNG/C** (donde N es una mezcla equimolar de los 4 nucleótidos), esto produce una mezcla de codones que contienen los 20 diferentes aminoácidos más un codón de término. Siguiendo la estrategia de mutagénesis descrita en el artículo 2, en la siguiente figura se resume las etapas de combinación de mutaciones dirigidas por los diferentes oligonucleótidos, después de la tercera etapa el oligo mutagénico 4 puede ser de dos tipos: mutante en la posición 73 y silvestre en la posición 70 ó mutante en la posición 73 y mutante (serina → cisteína) en la posición 70.

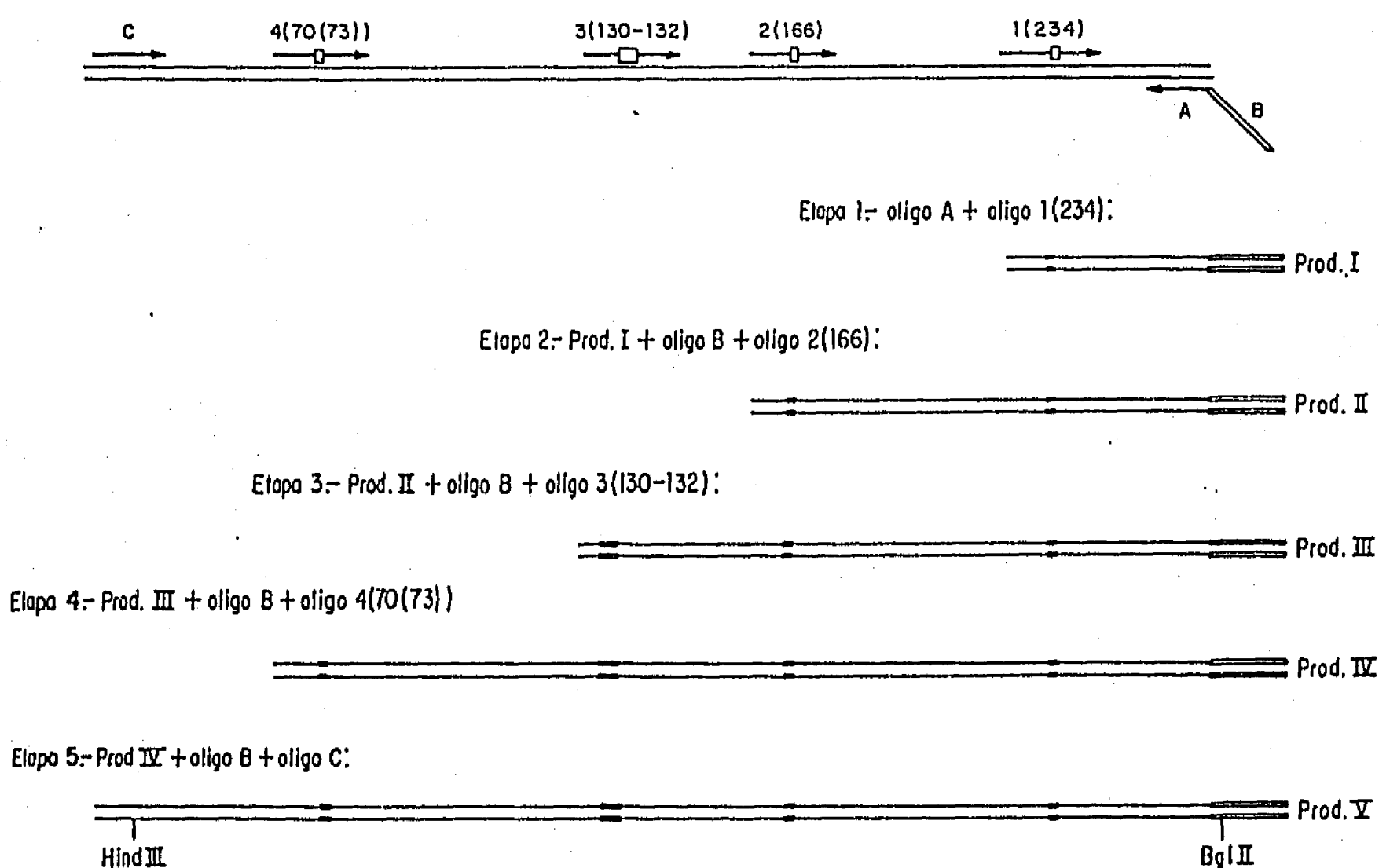


Fig. 4 Esquema de las diferentes etapas de la mutagénesis del sitio activo de la β -lactamasa. Las flechas indican la posición de los diferentes oligos para la construcción de los genes mutantes. Los productos de cada etapa son portadores de las mutaciones introducidas (zonas negras \blacksquare) y de la región necesaria para la amplificación selectiva de productos mutantes (el oligo B sólo es complementario a los productos). Cada etapa requiere de cuatro a cinco ciclos de PCR sólo con producto previo y templado, antes de agregar los oligos respectivos.

Lo anterior quiere decir que de la cuarta etapa en adelante se requiere de reacciones separadas para construcción de los genes mutantes con serina o cisteína catalítica. Por razones desconocidas no funcionó la quinta etapa de PCR y se recurrió a un producto IV ligeramente modificado. Este producto IV' consistió en una extensión de 18 pares de bases cercanas a la posición 70 lo cual se logró con una reacción adicional de PCR utilizando el siguiente oligonucleótido:

5'-GCCCCGAAGAACGTTTTCCAATGATG(A o T)GCACT-3', el producto IV original correspondiente como templado y el oligonucleótido B (tapón). Cabe mencionar que a cada producto de PCR se le sometió a una reacción de polimerización con el producto Klenow de la polimerasa en presencia de los 4 nucleótidos, lo cual fue necesario para eliminar un nucleótido 3' terminal que se agrega durante la reacción de PCR (46-47).

Clonación y electroporación del banco de genes mutantes.

Después de purificar los genes mutantes de la β -lactamasa, el producto de PCR se digirió con las enzimas de restricción *Bgl*II y *Hind*III y se clonó en el vector de selección positiva pKGS (48) digerido en los sitios *Bgl*II y *Hind*III del gene de la endonucleasa *Eco*RI. El número de transformantes analizados hasta la fecha (aprox. 5×10^4) es demasiado pequeño para un banco tan complejo de genes mutantes (20^6 secuencias diferentes de mutantes). Se tiene la secuencia nucleotídica de varias mutantes inactivas lo cual ha servido para comprobar la riqueza de mutantes en los sitios planeados. Dada la propensión a error de la Taq polimerasa se han observado mutaciones no planeadas en un número esperado según la tasa de error reportada para esta enzima.

Inesperadamente, de los sitios blanco de la mutagénesis, en la posición 234 no se ha observado la aparición de variantes. En los demás residuos, en ocasiones se ha observado la presencia del codón silvestre en alguno de ellos a pesar de existir modificación en otras posiciones. Hasta el momento desconocemos cual es el origen de la no variabilidad en esos aminoácidos, pero seguramente son errores del tipo de sistema de combinación de mutaciones.

Discusión

EcoRI

Nuestros resultados (artículo 1) y los de otros (49) claramente sugieren la existencia de una posible sobredeterminación de la especificidad de reconocimiento por la enzima *EcoRI*.

Los reportes preliminares de un nuevo estudio cristalográfico de un complejo *EcoRI*-DNA (23), permiten analizar nuestros resultados a la luz del papel real que pudieran jugar los 3 residuos que anteriormente se postulaban como los responsables de la especificidad de reconocimiento.

De las mutantes que nosotros obtuvimos, la variante **Glu144** → Cys fue la de mayor actividad enzimática según nuestro ensayo *in vivo*. En la nueva estructura del complejo, este residuo parece estar involucrado en una red de contactos entre varios aminoácidos de las dos subunidades que supuestamente ayuda a mantener la orientación correcta de las cadenas laterales de los aminoácidos que interaccionan directamente con las bases. Según nuestra variante, cisteína en la posición 144 pudiera ser capaz de mantener todavía algún tipo de interacción con esta red de contactos, sin embargo, algunas variantes en esta posición, como valina ó treonina, presentan actividad parcial a pesar del hecho de carecer totalmente de la capacidad de poder formar interacciones de puente de hidrógeno. **Arg145** parece mantener los contactos directos con las adeninas

del sitio de reconocimiento, además de estar también involucrada en la red de contactos mencionada anteriormente. De nuestras mutantes, solo lisina en esta posición produce una enzima con actividad parcial. Es interesante notar que la doble mutante que encontramos tolera treonina en la posición 145, aminoácido todavía capaz de formar puente de hidrógeno. **Arg200** está involucrada en mantener contactos indirectos con las bases del sitio de reconocimiento, mediados por una molécula de agua, y auxiliada por arginina en la posición 203. Cisteína en la posición 200 genera una variante con actividad parcial. Es interesante que en nuestras manos, el cambio a lisina en esta posición inactiva a la enzima, así como los cambios a histidina ó a asparagina, todos ellos residuos capaces de formar varios tipos de interacciones con el DNA.

Las generalizaciones que emergen del nuevo reporte de la estructura de la enzima en complejo con el DNA consisten en que la interfase proteína-DNA pudiera formarse cooperativamente y que los contactos de reconocimiento parecen ser redundantes, ésto significa que cualquier tipo de interacción pudiera ser removida sin perder la especificación por la base correcta, a pesar de que el nivel de discriminación pudiera ser reducido.

β -lactamasa

Los principales obstáculos que ha tenido el desarrollo de este proyecto han sido los siguientes:

1. transformación eficiente de la mezcla de genes mutantes.

Dada la complejidad de la mezcla posible de genes mutantes es necesario contar con una técnica de transformación con eficiencias de alrededor de 10^9 transformantes por microgramo de DNA. La electroporación es la técnica a elegir; sin embargo este tipo de enfoque exige DNA de alta pureza y en cantidades suficientes para lograr números de alrededor de 10^6 transformantes independientes. Todo lo anterior requiere de la optimización de cada etapa involucrada en la clonación de los genes mutantes al vector apropiado. El número examinado de transformantes hasta la fecha es tan pequeño que no permite la posibilidad de aislar clonas interesantes con baja representatividad en las secuencias posibles de mutantes.

2. contaminación con el gene silvestre.

Ya que nuestro interés consiste en la posibilidad de encontrar enzimas activas (con serina o cisteína catalítica) con un nuevo contexto de residuos adicionales que apoyen al residuo catalítico, requerimos de un ensayo selectivo que no tolere la mínima contaminación del gene silvestre. Durante el proceso selectivo hemos observado la aparición de clonas positivas que resultan ser silvestres. No sabemos cual es el origen de esta contaminación ya que durante las etapas de PCR el proceso de amplificación selectiva parece estar funcionando a la perfección. Con el fin de contender con este problema, hemos decidido utilizar como templado no al gene silvestre sino a versiones mutantes del mismo que son incapaces de conferir resistencia a ampicilina a células que lo contengan. Esperamos que esta variación elimine este foco de contaminación con el fin de asegurar el proceso selectivo de mutantes activas.

De los datos de secuencia que hemos obtenido, se ha observado la aparición de mutaciones adicionales a las planeadas. Esto fué más evidente cuando la estrategia de construcción de genes mutantes carecía de la etapa de tratamiento con Klenow de los productos de PCR. Por la naturaleza misma de este proyecto, no nos preocupan los cambios adicionales a los planeados, sin embargo, dada la dependencia absoluta del paso selectivo de mutantes activas, sería altamente favorable eliminar la posibilidad de contaminación con el gene silvestre.

Conclusiones

En este trabajo demostramos la posibilidad de producir cambios simultáneos en diferentes residuos lejanos en la secuencia de aminoácidos pero cercanos en la estructura tridimensional de una proteína. El método de mutagénesis combinatoria que utiliza PCR es robusto y rápido de implementarse. El empleo del oligonucleotido "tapón" es esencial para seleccionar la amplificación de sólo secuencias mutantes. Hemos experimentado las dificultades que se tienen tanto de clonación como de contaminación cuando se trabaja con reacciones de PCR, por lo que sugerimos ser extremadamente cuidadosos con el diseño así como con la ejecución de los experimentos de PCR. Los resultados obtenidos con la enzima *EcoRI* apoyan las propuestas que indican la necesidad de contactos adicionales a los mencionados en el primer reporte de la estructura tridimensional del complejo enzima-DNA. Los resultados con la enzima β -lactamasa son preliminares, pero se está trabajando en la optimización de las etapas necesarias para la obtención de un banco suficientemente grande de mutantes que permita la posibilidad de encontrar variantes interesantes. Se están haciendo esfuerzos por eliminar la fuente de contaminación del gene silvestre, la presencia del cual es desastrosa para el desarrollo exitoso del proyecto.

Artículo 10

GENE 06073

Combinatorial mutagenesis of three major groove-contacting residues of *EcoRI*: single and double amino acid replacements retaining methyltransferase-sensitive activities

(DNA-protein interactions; enzyme specificity; polymerase chain reaction; protein engineering)

Joel Osuna, Humberto Flores and Xavier Soberón

Centro de Investigación sobre Ingeniería Genética y Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos 62271 (Mexico)

Received by F. Bolívar: 6 May 1991

Revised/Accepted: 20 May/23 July 1991

Received at publishers: 29 July 1991

SUMMARY

A library of mutant *ecoRIR* genes encoding *EcoRI* restriction endonuclease was generated using trinucleotide blocks and a combination of recombinant DNA procedures, including primer extension and the polymerase chain reaction. Codons corresponding to three amino acids (E¹⁴⁴, R¹⁴⁵ and R²⁰⁰), previously implicated in the specific recognition of the DNA substrate, were combinatorially mutated so as to generate a library that potentially contains all 20³ possible single, double and triple aa replacements, in a balanced distribution. Inspection of the phenotypes of *Escherichia coli* colonies bearing the mutant genes showed that several of them retained activities that were deleterious to the cells but were still protected by the *EcoRI* methyltransferase. These included new enzyme variants, including non-conservative single (Thr or Val for Glu¹⁴⁴) and double (Val for Glu¹⁴⁴ and Thr for Arg¹⁴⁵) replacements.

INTRODUCTION

In the study of the structure-function relationship of proteins, as well as for the generation of proteins with novel properties, methods for the site-directed mutagenesis of the corresponding genes have become an indispensable tool. There are several stages, conceptual and methodological, to be taken into account when designing a rational muta-

genesis strategy: first, even when there is a known three-dimensional structure for the protein, it is extremely difficult to predict which changes are necessary to achieve the desired property of the protein. This type of prediction is still harder in the event that more than one residue change is needed. Second, frequently the residues one wishes to alter are far away from each other in the primary sequence (although they may be close together in the active conformation of the protein), and the combinatorial alteration of such residues is highly desirable (Wells, 1990). Third, as the number of target residues increases, the use of codons as mutagenic units becomes important in order to obtain a large number of useful alleles within a manageable mutant population (Sirotkin, 1986). We propose here a method for combinatorial, saturation mutagenesis, with codons as mutagenic units.

In order to test a mutagenesis method incorporating the considerations mentioned above, we chose the restriction endonuclease *EcoRI* as a model protein. The first report on the structure of the protein-DNA complex, to 3Å resolution

Correspondence to: Dr. X. Soberón, CIIGB/UNAM, Apdo. Postal 510-3, Col. Miraval 62271, Cuernavaca, Morelos (Mexico)
Tel. (52-73) 172-399; Fax (52-73) 172-388.

Abbreviations: aa, amino acid(s); bp, base pair(s); *ecoRIR*, gene encoding *EcoRI*; *ecoRIIM*, gene encoding M-*EcoRI*; ENase, restriction endonuclease; EtdBr, ethidium bromide; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, 1000 bp; Km, kanamycin; LB, Luria-Bertani (medium); M-*EcoRI*, *EcoRI* MTase; MTase, methyltransferase; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; ^R, resistance; ss, single-strand(ed); wt, wild type.

(McClarín et al., 1986), led to the proposition that three aa (E^{144} , R^{145} and R^{200}) directly contact the purine bases at the major groove of the substrate DNA molecule. A recent revision of the structure (Kim et al., 1990) reveals additional regions of the protein that contact the pyrimidine bases, also at the major groove, in a new model which involves differences in chain connectivity. The lethal activity of the enzyme *in vivo*, together with the existence of a corresponding MTase, which inhibits DNA cleavage by the ENase, has permitted the design of sufficiently sensitive *in vivo* assays for the detection of mutant ENases with either normal or altered activity and/or specificity for the canonical site (Heitman and Model, 1990; Oelgeschläger et al., 1990).

In this paper, we report the generation of a library of DNA fragments, potentially containing all of the 8000 possible aa combinations at three positions that face the DNA major groove (E^{144} , R^{145} and R^{200}). We used trinucleotide blocks in the oligo synthesis and obtained a fragment library which resulted in the expected distribution of replacements (single, double and triple), with no obvious bias and no stop codons. Using this library and a colony phenotype assay, we have isolated new partially active mutants, including a double aa replacement, that further strengthen the notion of the importance of DNA conformation (Lesser et al., 1990), and/or additional contacts at the bases (Heitman and Model, 1990), on the recognition process.

RESULTS AND DISCUSSION

(a) Combinatorial mutagenesis of aa 144, 145 and 200

In order to have access to all 20^3 possible combinations of aa at these three positions, we designed an experiment in which we would introduce, via synthetic DNA, a fragment consisting of 50% wt codon, 50% of an equimolar mixture of codons for the 20 aa, at each of the three positions (see Fig. 1 legend). In the ideal case (e.g., with equal coupling efficiencies of the trinucleotides and no other artifacts during chemical and enzymatic manipulations), a binomial distribution would predict the composition of the fragment to be 12.5% wt in all three positions, 37.5% single, 37.5% double and 12.5% triple replacements. The number of clones sequenced so far (as shown in Table I) is still too low to tell how close the actual distribution of changes is, compared to the ideal. Nonetheless, our results do show that we have a collection with a variety of replacements, in all three positions, which has already provided mutants with interesting phenotypes (see section b).

This mutagenesis scheme goes beyond previous reports on combinatorial mutagenesis (Reidhaar-Olson and Sauer, 1988; Dunn et al., 1988; Sartorius et al., 1989). First, we have targeted residues that are far away from each other in

the sequence. Second, through the use of trinucleotides we have aimed at a non-biased collection of replacements,

TABLE I

Variant enzyme phenotypes

<i>EcoRI</i> mutants ^a	aa ^b			Plating phenotype ^c without M· <i>EcoRI</i> ^d	
	E^{144}	R^{145}	R^{200}	-IPTG	+IPTG
	wt	—	—	—	+
E144C	C	—	—	++	—
E144T	T	—	—	++++	+
E144V	V	—	—	++++	+
E144G	G	—	—	++++	++++
E144F	F	—	—	++++	++++
E144W	W	—	—	++++	++++
R145K	—	K	—	++++	+++
R200C	—	—	C	+++	+
R200K	—	—	K	++++	++++
R200N	—	—	N	++++	++++
R200H	—	—	H	++++	++++
R200Q ^e	—	—	Q	ND	ND
ER-VT ^f	V	T	—	++++	++
ER-KY	K	Y	—	++++	++++
ER-PF	P	F	—	++++	++++
ER-YW ^e (ASN 149)	Y	W	—	ND	ND
ER-DE ^e	D	—	E	ND	ND
ERR-VTG	V	T	G	++++	++++
ERR-JKL (LEU 203)	I	K	—	++++	++++
ERR-DIH ^e	D	I	H	ND	ND
ERR-RHT	R	H	T	++++	++++

^a Designations for variant ENase indicate the original aa at the left and the aa present in the mutant, at the right, separated by the aa number where the replacement occurred. Numbers for multiple replacements are omitted for simplicity, they can be derived from columns 2–4. Numbers in parentheses denote mutations present outside the intended mutagenesis window.

^b Sequencing was performed on the entirety of *PstI-HindIII* fragments, derived from mutant colonies, cloned on M13mp19. The same fragments were also recloned in *PstI-HindIII*-digested pKGS to verify the observed phenotypes (shown in Fig. 2).

^c JM101 colony phenotypes were assessed visually. Colonies were grown on Km plates at a final concentration of 50 μ g/ml. Where indicated, plates were supplemented with IPTG to a final concentration of 1 mM. + + + + +, normal appearing colony; + + + +, lower density colonies but similar size; + + + and + +, reduced size and visibly translucent colonies; +, flat, small and translucent colonies (similar appearance as the parent enzyme). —, no growth. ND, not determined.

^d When the mutant genes were expressed in a MTase context, all colonies had a normal phenotype, including colonies bearing the parent *ecoRIR* gene. Under induction conditions only the parent gene is lethal (see Fig. 2).

^e Data obtained from direct cloning of the *BglII-PstI* fragment in M13 vectors.

^f In this case, the *ecoRIM* gene was complemented both *in trans*, in a compatible plasmid, and *in cis*, with the *ecoRIM* gene cloned in the same plasmid as the *ecoRIR* gene. In both cases, there was a protective effect of the MTase on the lethality of the mutant ENase under induction conditions. The rescue was more evident in the cells containing the *ecoRIM* gene present *in trans* (see Fig. 2).

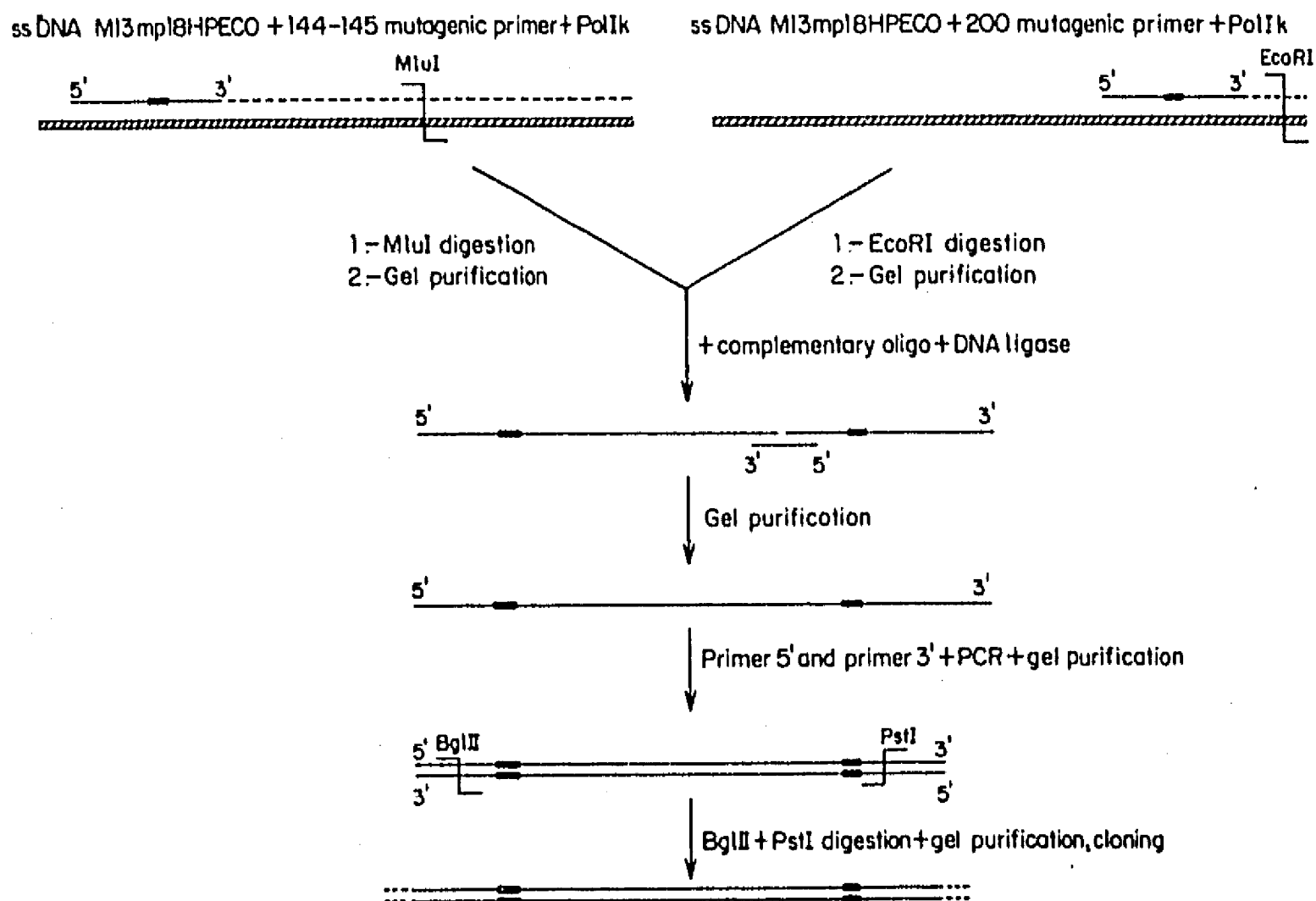


Fig. 1. Combinatorial mutagenesis strategy. The short, solid lines indicate synthetic oligos used as mutagenic primers or as ligation adapters (thickenings denote mutagenic regions). Hatched double lines denote the M13mp18 ss DNA template. Dashed lines depict the extension products. Methods: All cloning experiments were carried out with M13mp18, M13mp19 (Yanisch-Perron et al., 1985) or pKGS (Kuhn et al., 1986) as vector DNA and JM101 (Messing, 1979) as recipient strain. Plasmid pKGS, which carries a full length *ecoRIR* gene under the control of the inducible promoter lacUV5, and pSC101meth, a Tc^R derivative of pSC101, containing the *ecoRIM* gene (Betlach et al., 1976), were a generous gift from P.J. Greene (UCSF). Oligos were synthesized on a DNA synthesizer (Systec 1460A) using the phosphoroamidite method (Beaucage and Caruthers, 1981) and controlled pore glass as support (Adams et al., 1983). In some steps, those involving the introduction of a cocktail of trinucleotides, a mixture was prepared containing equimolar amounts of 20 different, fully protected, phosphotriester trinucleotides, synthesized as described (Broka et al., 1980). Each trinucleotide corresponded to a codon for one aa, as follows: Ala, GCT; Arg, CGT; Asn, AAC; Asp, GAC; Cys, TGT; Glu, GAA; Gln, CAG; Gly, GGT; His, CAC; Ile, ATC; Leu, CTG; Lys, AAA; Met, ATG; Phe, TTT; Pro, CCG; Ser, TCT; Thr, ACT; Trp, TGG; Tyr, TAC and Val, GTA. At the positions corresponding to codons selected for replacements, the growing oligo was reacted, using the procedure of Ito et al. (1982), through a manual injection valve, with a solution made with a total of 20 μmol trinucleotides from the cocktail described above and 20 μmol of the trinucleotide corresponding to the aa present in the wt *EcoRI* sequence. Purification of oligos was performed by gel electrophoresis in 20% polyacrylamide-8 M urea gels. Large pieces were sliced from the gel to allow for heterogeneity of migration of the mixed oligos. For the creation of convenient restriction sites, oligo-directed mutagenesis was performed essentially as described (Su and El-Gewely, 1988). M13mp19HPECO, an M13mp19 clone that carries the *HindIII-PstI* insert, which corresponds to the fragment coding for aa 68 to 206 of the ENase was employed as a single stranded template. We simultaneously added all three mutagenic primers 5'-GCCATTAGATCTTGATCTCCTC, 5'-GACCCCTCTAGAAAAGGACGT, 5'-TTAACAACGCGTCCATCTGGTC, plus the M13 'universal' primer. These oligos introduced silent mutations creating *BglII*, *XbaI* and *MluI* sites, at positions corresponding to aa 135, 169 and 187, respectively. A clone scoring positive for all three mutations was chosen and, after sequence verification, used as a source of DNA to return the *HindIII-PstI* fragment, bearing the mutations, to the *ecoRIR* gene of pKGS plasmid, substituting for the original region. Combinatorial, site saturation mutagenesis at aa positions 144, 145 and 200 of *EcoRI* endonuclease was approached by the following strategy: M13mp18HPECO, an M13mp18 clone containing the *HindIII-PstI* insert of the *ecoRIR* gene was used to infect *E. coli* strain JM101. The ss DNA was purified and used as template for primer extension using the two mutagenic oligos (5'-ATCAAGATCTAATGGCTGCTGGTAATGCTATCXXXXTCTCATAAGA, 5'-CGCGTTGTTAATCTTGAGTATAAT-TCTGGTATATTAATXXXTTAGATCGAC, where the X denote positions mutagenized in the manner described above) in separate experiments. The mutagenic oligos were phosphorylated by T4 polynucleotide kinase using [γ - ^{32}P]ATP. The radioactive oligos were annealed to the ss DNA and treated with PolIk for 1 h at 37°C. After inactivating the polymerase by heating at 65°C for 10 min, the DNA was digested with *MluI* or *EcoRI* depending on the oligo used in the extension. The ss DNA fragment produced was purified by 20% polyacrylamide-8 M urea gel after visualization of the band by autoradiographic exposure. The ss DNA fragments of 161 and 99 nt (corresponding to the expected lengths of the fragments generated with the oligos complementary to the regions of aa 144-145 and 200, respectively) were gel purified. The fragments (approx. 0.03 pmol each) were then ligated to one another, aided by 0.3 pmol of an adaptor oligo (5'-TTAACAACGCGTCCATCTGGTC) in a 20 μl reaction. The expected 260 nt ss DNA product was gel purified and subsequently amplified by PCR. We used 5'-ATCAAGATCTAATGGCTGCTGGTAA and 5'-CTAGAGTCGACCTGCAGTTA as primers, and performed 30 cycles at 95, 55 and 65°C, for 1.5, 1.5 and 3 min, respectively. The amplified fragment was ligated to adapters (due to the unexpected difficulty to digest with *BglII* at the site located 4 bp from the end of the fragment) and then digested with *BglII* + *PstI*. The purified fragment was cloned either in pKGS, replacing the wt with the mutant region as a cassette, or in M13mp19, for direct sequencing.

containing a suitable distribution of triple, double and single replacements, with no stop codons; with methods based on synthesis with mononucleotides, the resultant mutant library would be expected to be poorer, due to inherent limitations imposed by the degeneracy of the genetic code as well as the fact that the triplet is the coding unit. As the advantage of using triplets for mutagenesis, although significant (Sirotkin, 1986), is probably offset by the general difficulty of producing and using such synthetic units, we are currently working at further advancing these methods, adapted to more widely available synthetic chemistries. Our efforts are also stimulated by the versatility attainable by the use of PCR.

(b) Phenotypes of bacteria bearing the mutant *ecoRIR* genes

The transformation mixture of the mutant fragment cloned in pKGS was plated in conditions in which the gene is repressed (no IPTG added). The *ecoRIR* gene present in this plasmid carries a mutation that results in a protein product, altered at one aa ($E^{160} \rightarrow D$) (Kuhn et al., 1986), and with lower activity under repressing conditions (note that all of the mutants reported in the present study contain this replacement as well). The presence of this variant of the *ecoRIR* gene, under the repressed conditions, is tolerated but confers a translucent, flat appearance to the *E. coli* JM101 colony bearing it. With the mutant collection, we observed colonies with at least three distinctly different phenotypes: normal, intermediate, and wholly flat-translucent (Fig. 2). Furthermore, when these colonies were replica-plated to media containing the inducer, we could

again observe differences among them, namely: normal growth, poor growth and no growth (see Fig. 2 and Table I). To verify the correspondence between the sequence and the observed phenotype, plasmid DNA was purified and used both for retransformation and as a source for the sequencing.

Out of the limited number of colonies analyzed so far (about 10^3 , from several transformation experiments), some preliminary observations could be noted. Some 10% of the colonies displayed phenotypes indistinguishable from those carrying the parent gene, under repressing conditions. Sequencing experiments from several of such colonies revealed the presence of the expected silent mutations (coming from codons present in the trinucleotide mixture used to synthesize the oligo, which are synonymous, but different from the original).

Less than 1% of the colonies showed the intermediate phenotype; sequencing of several such clones revealed only the Cys^{144} and Cys^{200} variants. From a few hundred colonies that were replica plated to inducing conditions, about 2% manifested a distinguishably poorer growth ($E^{144} \rightarrow V$, $E^{144} \rightarrow T$, $R^{145} \rightarrow K$, and $E^{144}R^{145} \rightarrow VT$ were identified this way). Overall, therefore, there is an indication that less than 5% of the variant *EcoRI* altered in one or more of these three residues confers a phenotype to the cells carrying them (as detected in our assay system).

Sensitivity to the presence of the $M \cdot EcoRI$ was tested for some of the mutants either by transforming the corresponding plasmid DNA into JM101 cells bearing the *ecoRIM* gene in a compatible plasmid or by replacing the mutant fragment in a plasmid that contains both the *ecoRIR*

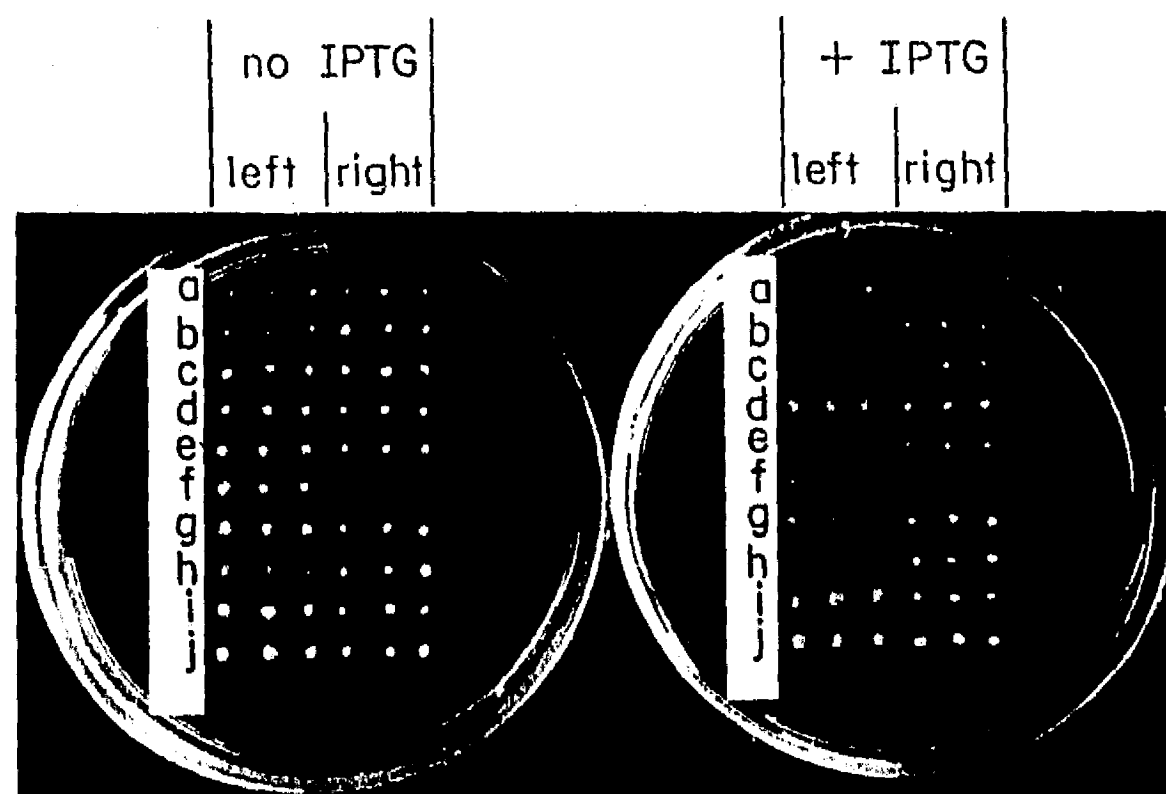


Fig. 2. Colony phenotypes of variant ENases. Colonies expressing *EcoRI* mutant ENases are as follows: a, parent enzyme; b, $E^{144} \rightarrow C$; c, $E^{144} \rightarrow T$; d, $E^{144} \rightarrow G$; e, $E^{144}R^{145} \rightarrow VT$; f, $E^{144}R^{145} \rightarrow VT + M \cdot EcoRI$ (present in *cis*, see Table I); g, $R^{145} \rightarrow K$; h, $R^{200} \rightarrow C$; i, $R^{200} \rightarrow K$; and j, $R^{200} \rightarrow N$. Left plate, no IPTG inducer, right plate, 1 mM IPTG. Each plate is divided in two columns containing three independently isolated colonies from a fresh transformation. Colonies in the left and right columns come from host strains lacking and containing the $M \cdot EcoRI$, respectively.

and the *ecoRIM* genes. All mutants tested showed a MTase-sensitive phenotype in one or more conditions (Fig. 2). It was striking to observe variants with remanent activity that are still MTase-sensitive, especially those containing entirely non-conservative (e.g., E¹⁴⁴ → V) as well as double (E¹⁴⁴ → V + R¹⁴⁵ → T) aa replacements.

Our results are in general agreement with recent reports (Heitman and Model, 1990; Needels et al., 1989; Alves et al., 1989) showing that some replacements at aa 144, 145 and 200 of *EcoRI* generate ENases with reduced activities, but are still specific towards the GAATTC sequence.

The use of combinatorial mutagenesis, together with our assay system and, possibly, the utilization of the mutant at position 160, permitted the identification of the double replacement variant that retains activity against the canonical site. It is noteworthy that the phenotypes conferred by some of the mutations we isolated differ from those previously reported. For instance, mutants E¹⁴⁴ → T and E¹⁴⁴ → V have detectable in vivo activities in our hands, whereas Heitman and Model (1990) listed them as inactive. On the other hand, we observed a null phenotype for E¹⁴⁴ → G, listed as marginally active by the same authors. Since our detection system is different and uses a different strain from those previously reported, it is possible that mutations that result in mechanistically different activities (e.g., nicking, dissociation prior to hydrolysis of the second strand, etc.) stand out in one assay, but not in the other. We are currently working on the isolation of the mutations at aa 144, 145 and 200 from the lesion at position 160, which is also a conceivable cause of the discrepancies noted above.

Due to the amendments on the crystal structure of *EcoRI*, there has been caution in interpreting recent studies of *EcoRI* in structural terms (Lesser et al., 1990). Based on the description from the article reporting the revised, refined, coordinate set (Kim et al., 1990), as well as on our own observations using the available α -carbon coordinates (Brookhaven PDB entry 1RIE), we infer that the side chains of aa 145 and 200 are within hydrogen bonding distance of the bases at the major groove of the DNA molecule, with aa 144 also nearby. Therefore, although our results substantiate the notion that additional contacts are involved in the determination of specificity, they clearly pertain to the protein-DNA interface.

(c) Conclusions

In summary, we have implemented a method to combinatorially mutagenize three residues implicated in the specific recognition of DNA by endonuclease *EcoRI* (McClarín et al., 1986). Our collection is composed of a balanced proportion of single, double and triple replacements and allowed us to isolate several mutants that retain activity. The observation of an increasing number of altered proteins with replacements at these three residues, that still retain

activity and specificity, reinforces the experimental support of the notion that other aa play a significant role in the recognition and cleavage processes. Work under way is aimed at the screening of a sufficiently large number of clones in order to score the phenotypes of other multiple replacements as well as extending the mutagenesis window. We believe that our approach, applied to all residues involved in hydrogen bonding interactions, as interpreted from the new structure, should be valuable in providing information with regard to the possible overdetermination of the specificity of this enzyme.

ACKNOWLEDGEMENTS

We are grateful to Paul Gaytán for oligo synthesis, Maria Elena Munguía for technical assistance, Sonia Caro for secretarial help and Mario Zurita for critically reading the manuscript. J.O. is supported by a Ph.D. scholarship from Consejo Nacional de Ciencia y Tecnología, México.

REFERENCES

- Adams, S.P., Kauka, K.S., Wykes, E.J., Holder, S.B. and Gallupi, G.R.: Hindered dialkyl amino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. *J. Am. Chem. Soc.* 105 (1983) 661-663.
- Alves, J., Rüter, T., Geiger, R., Fliess, A., Maass, G. and Pingoud, A.: Changing the hydrogen-bonding potential in the DNA binding site of *EcoRI* by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site -GAATTC-. *Biochemistry* 28 (1989) 2678-2684.
- Beaucage, S.L. and Caruthers, M.H.: Deoxynucleoside phosphoramidites. A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* 22 (1981) 1859-1862.
- Betlach, M., Hershfield, V., Chow, L., Brown, W., Goodman, H.M. and Boyer, H.W.: A restriction endonuclease analysis of the bacterial plasmid controlling the *EcoRI* restriction and modification of DNA. *Fed. Proc.* 35 (1976) 2037-2043.
- Broka, C., Hozumi, T., Arentzen, R. and Itakura, K.: Simplifications in the synthesis of short oligonucleotide blocks. *Nucleic Acids Res.* 8 (1980) 5461-5473.
- Dunn, I.S., Cowan, R. and Jennings, P.A.: Improved peptide function from random mutagenesis over short 'windows'. *Prot. Eng.* 2(4) (1988) 283-291.
- Heitman, J. and Model, P.: Substrate recognition by the *EcoRI* endonuclease. *Proteins* 7 (1990) 185-197.
- Ito, H., Ike, Y., Ikuta, S. and Itakura, K.: Solid phase synthesis of polynucleotides, VI. Further studies on polystyrene copolymers for the solid support. *Nucleic Acids Res.* 10 (1982) 1755-1769.
- Kim, Y., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M.: Refinement of *EcoRI* endonuclease crystal structure: a revised protein chain tracing. *Science* 249 (1990) 1307-1309.
- Kuhn, I., Stephenson, F.H., Boyer, H.W. and Greene, P.J.: Positive-selection vectors utilizing lethality of the *EcoRI* endonuclease. *Gene* 44 (1986) 253-263.
- Lesser, D.R., Kurpiewski, M.R. and Jen-Jacobson, L.: The energetic basis of specificity in the *EcoRI* endonuclease-DNA interaction. *Science* 250 (1990) 776-786.

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

- McClarín, J.A., Frederick, C.A., Wang, B., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M.: Structure of the DNA-*EcoRI* endonuclease recognition complex at 3 Å resolution. *Science* 234 (1986) 1526-1541.
- Messing, J.: A multipurpose cloning system based on the single stranded DNA bacteriophage M13. *Recombinant DNA Technical Bulletin*, NIH Publication No. 79-99 2(2) (1979) 43-48.
- Needels, M.C., Fried, S.R., Love, R., Rosenberg, J.M., Boyer, H.W. and Greene, P.J.: Determinants of *EcoRI* endonuclease sequence discrimination. *Proc. Natl. Acad. Sci. USA* 86 (1989) 3579-3583.
- Oelgeschläger, T., Geiger, R., Rüter, T., Alves, J., Fliess, A. and Pingoud, A.: Probing the function of individual amino acid residues in the DNA binding site of the *EcoRI* restriction endonuclease by analysing the toxicity of genetically engineered mutants. *Gene* 89 (1990) 19-27.
- Reidhaar-Olson, J.F. and Sauer, R.T.: Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241 (1988) 53-57.
- Sartorius, J., Lehming, N., Kisters, B., Von Wilcken-Bergmann, B. and Müller-Hill, B.: Lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. *EMBO J.* 8(4) (1989) 1265-1270.
- Sirotkin, K.: Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. *J. Theor. Biol.* 123 (1986) 261-279.
- Su, T.-Z. and El-Gewely, M.R.: A multisite-directed mutagenesis using T DNA polymerase: application for reconstructing a mammalian gene. *Gene* 69 (1988) 81-89.
- Wells, J.A.: Additivity of mutational effects in proteins. *Biochemistry* 29 (1990) 8509-8517.
- Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33 (1985) 103-119.

Artículo 2º

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

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

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

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

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

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

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

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

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

Dear Authors,
Please cite ref 5 in text. Thankx

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

Not familiar, could you use another word - Many

to M13mp19 digested with the same two enzymes and transformed into *E. coli* DH5 α TM.

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

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

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

phrase change - ok?


REFERENCES

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

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

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

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



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

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

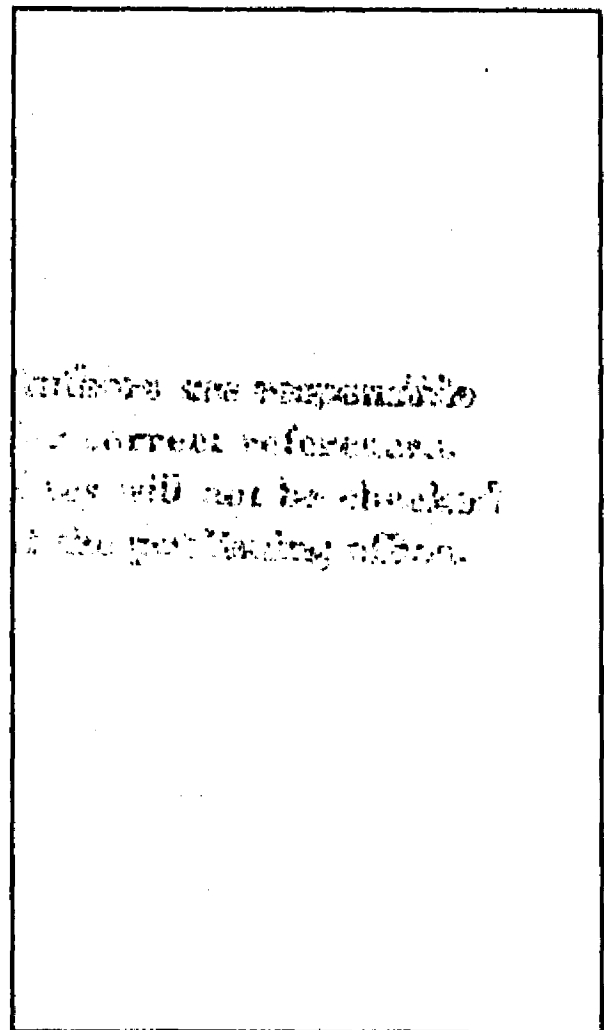
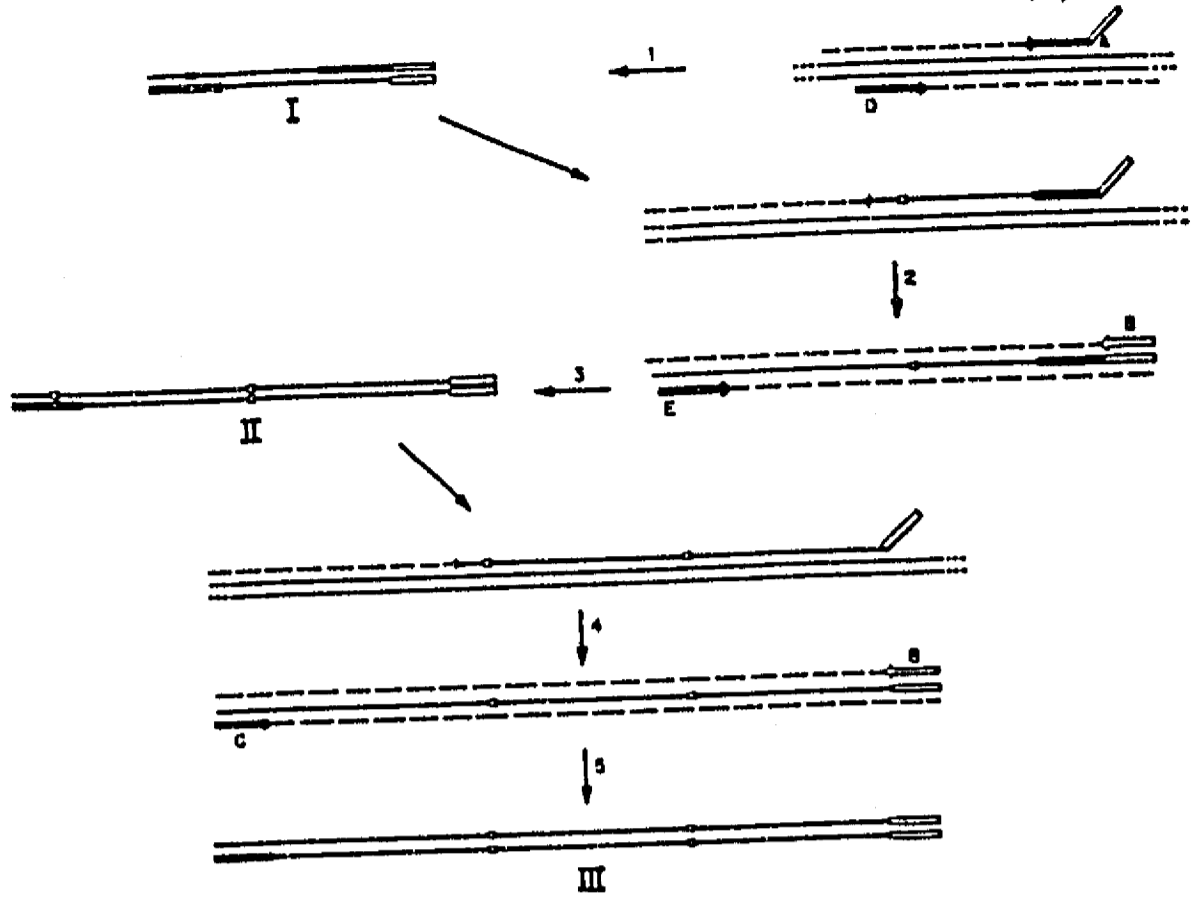


Figure 2. Electrophoretic analysis of the PCR products. A, size markers; B, C and D, PCR products I, II and III (see Figure 1), respectively.

F1
BT 91-202
Merino



F2
BT 91-202
Merino



Referencias

- 1• Wilson,C., Mace,J.E., and Agard,D.A. J. Mol. Biol. 220:495-506, (1991).
- 2• Lee,C., and Levitt,M. Nature 352:448-451, (1991).
- 3• Wells,J.A. Biochemistry 29:8509-8517, (1990).
- 4• Scrutton,N.S., Berry,A., and Perham,R.N. Nature 343:38-43, (1990).
- 5• Dunn,C.R., Wilks,H.M., Halsall,D.J., Atkinson,T., Clarke,A.R., Muirhead,H., and Holbrook,J.J. Phil. Trans. R. Soc. Lond. 332:176-184, (1991).
- 6• Meng,M., Lee,C., Bagdasarian,M., and Zeikus,J.G. Proc. Natl. Acad. Sci. 88:4015-4019, (1991).
- 7• Bradley,M., Bücheler,U.S., and Walsh,C.T. Biochemistry 30:6124-6127, (1991).
- 8• Henderson,G.B., Murgolo,N.J., Kuriyan,J., Osapay,K., Kominos,D., Berry,A., Scrutton,N.S., Hinchliffe,N.W., Perham,R.N., and Cerami,A. Proc. Natl. Acad. Sci. 88:8769-8773, (1991).
- 9• Khouri,H.E., Vernet,T., Ménard,R., Parlati,F., Laflame,P., Tessier,D.C., Gour-Salin,B., Thomas,D.Y., and Storer,A.C. Biochemistry 30:8929-8936, (1991).
- 10• Carter,P., and Wells,J.A. Proteins 7:335-342, (1990).
- 11• Evin,L.B., Vásquez,J.R., and Craik,C.S. Proc. Natl. Acad. Sci. 87:6659-6663, (1990).
- 12• Bass,S., Greene,R., and Wells,J.A. Proteins 8:309-314, (1990).
- 13• McCafferty,J., Griffiths,A.D., Winter,G., and Chiswell,D.J. Nature 348:552-554, (1990).
- 14• McCafferty,J., Jackson,R.H., and Chiswell,D.J. Protein Eng. 4:955-961, (1991).
- 15• Lim,W.A., and Sauer,R.T. J. Mol. Biol. 219:359-376, (1991).
- 16• York,J.D., Li,P., and Gardell,S.J. J. Biol. Chem. 266:8495-8500, (1991).
- 17• Wissmann,A., Wray,L.V., Sommaggio,U., Baumeister, R., Geissendörfer,M., and Hillen,W. Genetics 128:225-232, (1991).
- 18• Auld,D.S., and Pielak,G.J. Biochemistry 30:8684-8690, (1991).

- 19• Loeb,D.D., Swanstrom,R., Everitt,L., Manchester,M., Stamper,S.E., and Hutchison III,C.A. *Nature* 340:397-400, (1989).
- 20• Nelson,R.M., and Long,G.L. *Anal. Biochem.* 180:147-151, (1989).
- 21• McClarin,J.A., Frederick,C.A., Wang,B., Greene,P., Boyer,H.W., Grable,J., and Rosenberg,J.M. *Science* 234:1526-1541, (1986).
- 22• Kim,Y., Grable,J.C., Love,R., Greene,P., and Rosenberg,J.M. *Science* 249:1307-1309, (1990).
- 23• Rosenberg,J.M. *Curr. Opin. Struct. Biol.* 1:104-113, (1991).
- 24• Samraoui,B., Sutton,B.J., Todd,R.J., Artymiuk,P.J., Waley,S.G., and Phillips,D.C. *Nature* 320:378-380, (1986).
- 25• Herzberg,O., and Moulton,J. *Science* 236:694-701, (1987).
- 26• Dideberg,O., Charlier,P., Wéry,J.-P., Dehottay,P., Dusart,J., Ericum,T.,Frère,J.-M., and Ghuysen,J.-M. *Biochem. J.* 245:911-913, (1987).
- 27• Moews,P.C., Knox,J.R., Dideberg,O., Charlier,P., and Frère,J.-M. *Proteins* 7:156-171, (1990).
- 28• Ambler,R.P., Coulson,A.F.W., Frère,J.-M., Joris,B., Forsman,M., Levesque,R.C., Tiraby,G., and Waley,S.G. *Biochem. J.* 276:269-272, (1991).
- 29• Oliphant,A.R., and Struhl,K. *Proc. Natl. Acad. Sci.* 86:9094-9098, (1989).
- 30• Dube,D.K., and Loeb,L.A. *Biochemistry* 28:5703-5707, (1989).
- 31• Healey,W.J., Labgold,M.R., Richards,J.H. *Proteins* 6:275-283, (1989).
- 32• Herzberg,O. *J. Mol. Biol.* 217:701-719, (1991).
- 33• Escobar,W.A., Tan,A.K., and Fink,A.L. *Biochemistry* 30:10783-10787, (1991).
- 34• Adachi,H., Ohta,T., and Matsuzawa,H. *J. Biol. Chem.* 266:3186-3191, (1991).
- 35• Ellerby,L.M., Escobar,W.A., Fink,A.L., Mitchinson,C., and Wells,J.A. *Biochemistry* 29:5797-5806, (1990).
- 36• Delaire,M., Lenfant,F., Labia,R., and Masson,J.-M. *Protein Eng.* 4:805-810, (1991).
- 37• Jacob,F., Joris,B., Lepage,S., Dusart,J., and Frère,J.-M. *Biochem. J.* 271:399-406, (1990).

- 38• Lenfant,F., Labia,R., and Masson,J.-M. J. Biol. Chem. 266:17187-17194, (1991).
- 39• Brannigan,J., Matagne,A., Jacob,F., Damblon,C., Joris,B., Klein,D., Spratt,B.G., and Frère,J.-M. Biochem. J. 278:673-678, (1991).
- 40• Lamotte-Brasseur,J., Dive,G., Dideberg,O., Charlier,P., Frère,J.-M., and Ghuysen,J.-M. Biochem. J. 279:213-221, (1991).
- 41• Gibson,R.M., Christensen,H., and Waley,S.G. Biochem. J. 272:613-619, (1990).
- 42• Jacob,F., Joris,B., and Frère,J.-M. Biochem. J. 277:647-652, (1991).
- 43• Sigal,I.S., Harwood,B.G., and Arentzen,R. Proc. Natl. Acad. Sci. 79:7157-7160, (1982).
- 44• Bazan,J.F., and Fletterick,R.J. Proc. Natl. Acad. Sci. 85:7872-7876, (1988).
- 45• Sirotkin,K. J. Theor. Biol. 123:261-279, (1986).
- 46• Clark,J.M. Nucl. Acids Res. 16:9677-9686, (1988).
- 47• Mole,S.E., Iggo,R.D., and Lane,D.P. Nucl. Acids Res. 17:3319, (1989).
- 48• Kuhn,I., Stephenson,F.H., Boyer,H.W., and Greene,P.J. Gene 44:253-263, (1986).
- 49• Heitman,J., and Model,P. Proteins 7:185-197, (1990).