

OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

⑪ Número de publicación: **2 210 224**

⑤① Int. Cl.7: **C12N 15/12**, C12P 21/02
C07K 14/47, C12Q 1/68
G01N 33/68, G01N 33/53
A61K 48/00, A01K 67/027
C12P 21/08

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

- ⑧⑥ Número de solicitud europea: **90912428 .1**
⑧⑥ Fecha de presentación: **20.08.1990**
⑧⑦ Número de publicación de la solicitud: **0489058**
⑧⑦ Fecha de publicación de la solicitud: **10.06.1992**

⑤④ Título: **Gen de la fibrosis quística.**

③⑩ Prioridad: **22.08.1989 US 396894**
24.08.1989 US 399945
31.08.1989 US 401609

④⑤ Fecha de publicación de la mención BOPÍ:
01.07.2004

④⑤ Fecha de la publicación del folleto de la patente:
01.07.2004

⑦③ Titular/es:
HSC Research Development Corporation
555 University Avenue
Toronto Ontario M5G 1X8, CA
THE BOARD OF REGENTS ACTING FOR AND ON
BEHALF OF THE UNIVERSITY OF MICHIGAN

⑦② Inventor/es: **Tsui, Lap-Chee;**
Riordan, John, R.;
Collins, Francis, S.;
Rommens, Johanna, M.;
Iannuzzi, Michael, C.;
Kerem, Bat-Sheva;
Drumm, Mitchell, L. y
Buchwald, Manual

⑦④ Agente: **Tavira Montes-Jovellar, Antonio**

ES 2 210 224 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Gen de la fibrosis quística.

5 La presente invención se refiere, en general, al gen de la fibrosis quística (CF, del inglés cystic fibrosis) y, más particularmente, a la identificación, al aislamiento y a la clonación de la secuencia de ADN correspondiente a los genes de CF normal y mutante así como sus transcritos y productos génicos. La presente invención se refiere también a métodos para el rastreo y la detección de portadores de CF, la diagnosis de CF, el rastreo y diagnosis prenatal de CF y la terapia génica utilizando tecnologías recombinantes y la terapia de fármacos utilizando la información derivada del ADN, proteína y la función metabólica de la proteína.

15 La fibrosis quística (CF) es el trastorno genético recesivo autosomal grave más común en la población caucásica. Afecta aproximadamente a 1 de 2000 partos vivos en Norte América [Boat et al, The Metabolic Basis of Inherited Disease, 6ª ed., págs. 2649-2680, McGraw Hill, NY (1989)]. Aproximadamente 1 de cada 20 personas son portadores de la enfermedad.

20 Aunque la enfermedad se describió por primera vez a finales de los años 30, el defecto básico permanece siendo desconocido. Los síntomas principales de la fibrosis quística incluyen una enfermedad pulmonar crónica, insuficiencia exocrina pancreática y elevados niveles de electrolitos en el sudor. Los síntomas son consistentes con que la fibrosis quística es un trastorno exocrino. Aunque se han realizado avances recientes en el análisis del transporte de iones a través de la membrana apical del epitelio de células de pacientes de CF, no está claro que la regulación anormal de los canales cloruro represente el defecto principal en la enfermedad. Dada la carencia de comprensión del mecanismo molecular de la enfermedad, se ha realizado por lo tanto un enfoque alternativo en un intento de comprender la naturaleza del defecto molecular a través de la clonación directa del gen responsable sobre la base de su localización en el cromosoma.

30 Sin embargo, no existe ningún fenotipo claro que dirija un enfoque a la naturaleza exacta de la base genética de la enfermedad, o que permita una identificación del gen de la fibrosis quística. La naturaleza del defecto de CF en relación con los datos de la genética de población no ha sido fácilmente evidente. Tanto la prevalencia de la enfermedad como la heterogeneidad clínica han sido explicadas por varios mecanismos diferentes: elevada tasa de mutación, ventaja de heterocigotos, desviación genética, múltiples loci y compensación reproductora.

35 Muchas de las hipótesis no pueden ser ensayadas debido a la carencia de conocimiento del defecto básico. Por lo tanto, se han dirigido enfoques alternativos para la determinación y caracterización del gen de CF en un intento de identificar la localización del gen mediante análisis genéticos.

40 El análisis del enlace del gen de CF a los marcadores antigénico y de proteínas se intentó en los años 50, pero no se obtuvieron resultados positivos [Steinberg et al Am. J. Hum. Genet. 8: 162-176, (1956); Steinberg y Morton Am. J. Hum. Genet. 8: 177-189, (1956); Goodchild et al J. Med. Genet. 7: 417-419, 1976].

45 Más recientemente, ha resultado posible utilizar RFLPs para facilitar el análisis del enlace. El primer enlace de un marcador RFLP al gen de CF se describió en 1985 [Tsui et al. Science 230: 1054-1057, 1985], en el que el enlace se encontró entre el gen de CF y un marcador D0CRI-917 no caracterizado. La asociación se encontró en un análisis de 39 familias con niños afectados por CF. Esto demostró que, a pesar que no se había establecido la localización en el cromosoma, la localización del gen de la enfermedad se había reducido hasta aproximadamente el 1% del genoma humano, o aproximadamente 30 millones de pares de bases de nucleótidos.

50 La localización en el cromosoma de la sonda D0CRI-917 se estableció utilizando líneas de células híbridas de roedores-seres humanos que contenían diferentes complementos de cromosoma humano. Se demostró que D0CRI-917 (y, por lo tanto, el gen de CF) se correlaciona con el cromosoma 7 humano.

55 Se continuó con estudios de enlace físico y genético adicionales en un intento de detectar la localización del gen de CF. Zengerling et al [Am. J. Hum. Genet. 40: 228-236 (1987)] describen el uso de híbridos de células somáticas de seres humanos-ratón para obtener una relación física más detallada entre el gen de CF y los marcadores que se sabe se enlazan con el mismo. Esta publicación demuestra que el gen de CF puede asignarse a la región distal de la banda q22 o a la región proximal de la banda q31 en el cromosoma 7.

60 Rommens et al [Am. J. Hum. Genet. 43: 645-663, (1988)] dan una discusión detallada del aislamiento de muchas nuevas sondas 7q31. El enfoque esbozado condujo al aislamiento de dos nuevas sondas D7S122 y D7S340, que son próximas una a otra. La representación en el mapa de la electroforesis en gel de campo pulsado indica que estos dos marcadores RFLP se encuentran entre dos marcadores que se sabe flanquean el gen de CF, MET [White, R., Woodward S., Leppert M., et al. Nature 318: 382-384, (1985)] y D7S8 [Wainwright, B. J., Scambler, P. J., y J. Schmidtke, Nature 318: 384-385 (1985)], por lo tanto en la región del gen de CF. El descubrimiento de estos marcadores proporciona un punto de partida para el desplazamiento y salto de cromosomas.

65 Estivill et al, [Nature 326: 840-845 (1987)] describen que se localizó y caracterizó parcialmente un gen de ADNc candidato. Sin embargo, esto no enseña la localización correcta del gen de CF. La referencia describe un gen de ADNc candidato situado más abajo de una isla CpG, que son regiones ricas en nucleótidos GC submetiladas situadas más

arriba de muchos genes de vertebrados. La localización cromosómica del locus candidato se identifica como la región XV2C. Esta región se describe en la solicitud de patente europea 88303645.1. Sin embargo, esa región real no incluye el gen de CF.

5 Una dificultad principal en identificar el gen de CF ha sido la carencia de redistribuciones o deleciones cromosómicas detectables por citología, que facilitaron grandemente todos los éxitos previos de la clonación de genes de enfermedades humanas mediante el conocimiento de la posición en el mapa.

10 Redistribuciones y deleciones de este tipo pudieron observarse citológicamente y, como resultado, se podría correlacionar una localización física sobre un cromosoma particular con la enfermedad particular. Además, esta localización citológica se podría correlacionar con una localización molecular basada en la relación conocida entre sondas de ADN públicamente disponibles y alteraciones citológicamente visibles en los cromosomas. El conocimiento de la localización molecular del gen para una enfermedad particular permitiría la clonación y secuenciación de ese gen mediante procesos rutinarios, particularmente cuando el producto génico es conocido, y el éxito de la clonación se puede confirmar mediante el inmunoensayo de productos de expresión de los genes clonados.

15 En contraposición, en la técnica anterior no se conocía ni la localización citológica ni el producto génico del gen para la fibrosis quística. Con la reciente identificación de MET y D7S8, marcadores que flanqueaban el gen de CF pero que no detectaban su localización en la molécula, los autores de la presente invención proyectaron diversas nuevas estrategias para la clonación de genes para obtener un acercamiento al gen de CF de acuerdo con la presente invención. Los métodos empleados en estas estrategias incluyen el salto de cromosomas desde los marcadores flanqueantes, la clonación de fragmentos de ADN desde una región física definida con el uso de la electroforesis en gel de campo pulsado, una combinación de técnicas de clonación de híbridos de células somáticas y moleculares, diseñadas para aislar fragmentos de ADN procedentes de islas CpG submetiladas próximas a CF, microdissección y clonación de cromosomas, y clonación por saturación de un gran número de marcadores de ADN procedente de la región 7q31. Por medio de estas nuevas estrategias, los autores de la presente invención fueron capaces de identificar el gen responsable de la fibrosis quística en donde la técnica anterior era incierta o, incluso en un caso, incorrecta.

20 La aplicación de estas estrategias de clonación genética y molecular ha permitido el aislamiento y la clonación de ADNc del gen de la fibrosis quística sobre la base de su localización en el cromosoma, sin el beneficio de redistribuciones genómicas para detectar el camino. La identificación de las formas normal y mutante del gen y productos génicos de CF ha permitido el desarrollo de ensayos de rastreo y diagnóstico para CF utilizando sondas de ácidos nucleicos y anticuerpos contra el producto génico. A través de la interacción con el producto génico defectuoso y el recorrido en el que este producto génico está implicado, se hacen ahora posibles una terapia a través de la suplementación del producto génico normal y una manipulación y suministro de genes.

25 El gen implicado en el proceso de la enfermedad de la fibrosis quística, denominado en lo que sigue el "gen de CF" y sus equivalentes funcionales, ha sido identificado y aislado y clonado el ADNc, y sus transcritos y productos génicos han sido identificados y secuenciados. Se ha determinado que una delección de tres pares de bases que conduce a la omisión de un residuo fenilalanina en el producto génico corresponde a las mutaciones del gen de CF en aproximadamente un 70% de los pacientes afectados de CF, con diferentes mutaciones implicadas en la mayoría, si no en todos, los casos restantes.

30 Con la identificación y secuenciación del gen y de su producto génico, se pueden utilizar sondas de ácidos nucleicos y anticuerpos dirigidos contra el producto génico en una diversidad de ensayos de hibridación e inmunológicos para rastrear y detectar la presencia de un gen o producto génico de CF normal o uno defectuoso. También se pueden proporcionar estuches o kits de ensayo para un rastreo y diagnóstico de este tipo.

35 La terapia del paciente a través de la suplementación con el producto génico normal, cuya producción se puede amplificar utilizando técnicas genéticas y recombinantes, o su equivalente funcional, puede ser ahora posible. La corrección o modificación del producto génico defectuoso a través de medios de tratamiento con fármacos puede ser ahora posible. Además, la fibrosis quística se puede curar o controlar a través de una terapia génica corrigiendo el defecto del gen *in situ* o utilizando vehículos recombinantes u otros vehículos para suministrar una secuencia de ADN, capaz de la expresión del producto génico normal, a las células del paciente.

40 De acuerdo con un aspecto de la invención, se proporciona una molécula de ADN purificada que comprende un gen de fibrosis quística, comprendiendo dicho gen una secuencia de ADN seleccionada del grupo que consiste en:

- 45
- 50 (a) secuencias de ADN que corresponden a la secuencia de ADN de la figura 1 desde la posición 1 a la posición 1480 del residuo aminoácido o una forma polimórfica de dicha secuencia de ADN en la que la sustitución de nucleótidos de dicha secuencia de ADN no afecta a la función esencial del polipéptido codificado con ello;
 - 55 (b) secuencias de ADN que codifican un polipéptido regulador de la conductancia de la membrana de fibrosis quística (CFTR, del inglés cystic fibrosis transmembrane conductance regulator) normal que tiene la secuencia de acuerdo con la figura 1 desde la posición 1 a la posición 1480 de un residuo aminoácido o una forma polimórfica de dicho polipéptido, en el que la sustitución de aminoácidos en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta al funcionamiento esencial de la misma o su perfil hidropático o estructura secundaria o terciaria;

ES 2 210 224 T3

(c) secuencias de ADN que corresponden a un fragmento de la secuencia de ADN de la figura 1 que incluyen al menos 16 nucleótidos secuenciales entre las posiciones 134 y 4573 de la secuencia de nucleótidos y que pueden rastrear y detectar la presencia de un gen o producto génico de la fibrosis quística normal o defectuoso;

5 (d) secuencias de ADN que comprenden al menos 16 nucleótidos y que codifican un fragmento de la secuencia de aminoácidos de la figura 1 que pueden rastrear y detectar la presencia de un gen o producto génico de la fibrosis quística normal o defectuoso; y

10 (e) secuencias de ADN que codifican un epítipo codificado por al menos 18 nucleótidos secuenciales en la secuencia de ADN de la figura 1 entre las posiciones 1 y 1480 del residuo aminoácido.

El ADN puede ser, por ejemplo, ADNc.

15 De acuerdo con otro aspecto de la presente invención, se proporciona una molécula de ADN purificada de (a), (b) o (c) según se ha definido antes, que contiene, además, al menos una mutación en la secuencia de ADN que, si se expresa en células del cuerpo humano, se asocia con la función celular alterada que se correlaciona con la fibrosis quística, en donde dicha mutación incluye una deleción de tres nucleótidos que codifican fenilalanina en la posición 508 del residuo aminoácido.

20 De acuerdo con otro aspecto de la invención, se proporciona una sonda de ácido nucleico purificado que comprende una secuencia de nucleótidos de ADN o ARN que corresponde a la secuencia de acuerdo con las características (c), (d) o (e) según se define antes.

25 De acuerdo con otro aspecto de la invención, se proporciona un vector de clonación recombinante que comprende una molécula de ADN de acuerdo con la invención. El vector, de acuerdo con un aspecto de esta invención, está operativamente enlazado a una secuencia de control de la expresión en la molécula de ADN recombinante, de manera que se puede expresar la proteína CFTR normal o, alternativamente, con la otra secuencia de ADN mutante seleccionada se puede expresar el polipéptido CFTR mutante. La secuencia del control de la expresión se selecciona del grupo que consiste en secuencias que controlan la expresión de genes de células procarióticas o eucarióticas y sus virus y combinaciones.

En otro aspecto, la invención comprende un huésped no humano, transformado con un vector de la invención.

35 De acuerdo con otro aspecto de la invención, se proporciona un método para producir un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) normal o mutante, que comprende las etapas de:

(a) cultivar una célula huésped transfectada con el vector de la invención en un medio y en condiciones favorables para la expresión de un polipéptido CFTR normal o mutante; y

40 (b) aislar el polipéptido CFTR normal o mutante expresado.

De acuerdo con otro aspecto de la invención, se proporciona una molécula de ADN purificada que comprende una secuencia de ADN que codifica un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) mutante que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del residuo aminoácido, o una forma polimórfica de dicho polipéptido en la que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundario o terciaria, caracterizado además por una deleción de tres nucleótidos que codifican fenilalanina en la posición 508 del residuo aminoácido.

50 De acuerdo con otro aspecto de la invención, se proporciona un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) purificado, caracterizado por tener un peso molecular del péptido de aproximadamente 170.000 dalton, seleccionado del grupo que consiste en:

55 (a) un polipéptido CFTR normal que tiene actividad que afecta a la conductancia de iones de la transmembrana de la célula y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido o una forma polimórfica de dicho polipéptido CFTR normal, en que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria; y

60 (b) un polipéptido CFTR mutante que tiene actividad de fibrosis quística en células humanas y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido, excepto por la deleción de fenilalanina en la posición 508 del residuo aminoácido.

65 De acuerdo con otro aspecto de la invención, se proporciona un polipéptido codificado por la expresión de una secuencia de ADN de acuerdo con la invención, exhibiendo dicho polipéptido la actividad inmunológica o biológica de polipéptido CFTR normal o mutante.

ES 2 210 224 T3

De acuerdo con otro aspecto de la invención, se proporciona un procedimiento para aislar la proteína CFTR normal o mutante de acuerdo con la invención a partir de células que contienen dicha proteína, que comprende las etapas de:

- (a) solubilizar la proteína de una membrana celular seleccionada en la que se expresa dicha proteína CFTR normal o mutante, para proporcionar una solución de dicha proteína CFTR;
- (b) separar dicha proteína CFTR de dicha solución poniendo en contacto dicha solución con anticuerpos contra dicha proteína CFTR normal o mutante, estando dichos anticuerpos inmovilizados sobre un sustrato,
- (c) aclarar dicho sustrato para separar proteína no adherida a dichos anticuerpos;
- (d) liberar dicha proteína CFTR de dichos anticuerpos para aislar con ello dicha proteína CFTR, y
- (e) purificar dicha proteína CFTR para separar cualquier otra proteína de mamífero remanente.

De acuerdo con otro aspecto de la invención, un método de rastrear a un sujeto para determinar si dicho sujeto es un portador de fibrosis quística (CF) o un paciente de CF comprende:

proporcionar un ensayo para detectar en una muestra biológica a rastrear la presencia de al menos un miembro procedente del grupo que consiste en un gen de CF normal, productos génicos de CF normal, un gen de CF mutante, productos génicos de CF mutantes y sus mezclas, un polipéptido CFTR normal con una actividad que afecta a la conductancia de iones en la transmembrana de la célula y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido o una forma polimórfica de dicho polipéptido CFTR normal en que la sustitución de aminoácidos en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria, un polipéptido CFTR mutante con actividad de fibrosis quística en células humanas y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 de aminoácidos excepto la delección de fenilalanina en la posición 508 del residuo aminoácido, en donde el ensayo es

- (i) un ensayo que implica hibridación, o
- (ii) un ensayo inmunológico,

y/o se consigue utilizando una o más hibridaciones que utilizan oligonucleótidos específicos, secuenciación de ADN directa, enzimas de restricción, discriminación sobre la base de la movilidad electroforética en geles con agente desnaturizante, protección de RNasa, escisión química o el proceso de detección mediado por ligasa.

El ensayo de hibridación puede incluir el uso de uno o más nucleótidos de acuerdo con la invención, en donde dos nucleótidos no son los mismos, nucleótidos que pueden ser marcados por cualquier medio conveniente.

En los casos en los que el método de ensayo es inmunológico y la muestra biológica incluye un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) normal del sujeto y/o un polipéptido CFTR mutante del sujeto: teniendo dicho polipéptido CFTR normal la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido o una forma polimórfica de dicho polipéptido en la que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria; y teniendo dicho polipéptido CFTR mutante la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido, excepto la delección de fenilalanina en la posición 508 del residuo aminoácido.

El método de ensayo inmunológico puede incluir anticuerpos específicos para un polipéptido CFTR normal y/o mutante. Los anticuerpos para uso en el método pueden ser monoclonales.

De acuerdo con otro aspecto de la invención, el método de rastreo de un sujeto de la invención se lleva a cabo en un feto en el útero.

De acuerdo con otro aspecto de la invención, se proporciona un procedimiento para rastrear un portador o paciente de fibrosis quística (CF) potencial para identificar la presencia de una mutación de la fibrosis quística identificada en el gen de CF que codifica un polipéptido regulador de la conductancia en la transmembrana de la fibrosis quística (CFTR) que tiene la secuencia de acuerdo con la figura 1 desde la posición 1 a la posición 1480 del aminoácido o una forma polimórfica de dicho polipéptido, en cuya sustitución de aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria, incluyendo dicho procedimiento las etapas de:

- (a) hibridar una sonda de ADN de acuerdo con la reivindicación 3 sobre ADN genómico aislado procedente de dicho portador de CF potencial o dicho paciente potencial, abarcando dicha sonda de ADN dicha mutación de fibrosis quística en dicho gen de CF, en donde dicha sonda de ADN es capaz de detectar dicha mutación de la fibrosis quística; y

ES 2 210 224 T3

- (b) tratar dicho ADN genómico para determinar la presencia o ausencia de dicha sonda de ADN y, con ello, indicar, de acuerdo con una manera predeterminada de hibridación, la presencia o ausencia de dicha mutación de fibrosis quística.

5 En un aspecto adicional, la invención comprende un procedimiento para detectar portadores o pacientes de fibrosis quística, en donde dicho procedimiento consiste en poner en contacto un gen de CF del paciente o portador con una endonucleasa de restricción, determinar la presencia o ausencia de un sitio de endonucleasa de restricción en el gen de CF y comparar el modelo de los sitios de endonucleasa de restricción en el gen de CF con un gen de CF normal que corresponde a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1, la presencia de sitios adicionales al gen de CF normal o la ausencia de sitios en el gen de CF normal indicativos de un gen de CF mutante.

15 En otro aspecto de la invención, se proporciona un procedimiento para detectar portadores de fibrosis quística, en donde dicho procedimiento consiste en determinar una movilidad diferencial de productos de PCR heteroduplex en geles de poliacrilamida como resultado de inserciones o deleciones en el gen de CF mutante comparado con el gen de CF normal correspondiente a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1.

20 De acuerdo con otro aspecto de la invención, se proporciona un estuche para someter a ensayo, mediante inmunoensayo, la presencia de un gen de CF normal correspondiente a la secuencia de nucleótidos que codifica al polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1 o un gen de CF mutante que tiene al menos una deleción de tres pares de bases en el gen de CF normal, deleción que da como resultado la deleción de una fenilalanina procedente de un residuo aminoácido de la posición 508 de la figura 1, que comprende:

- 25 (a) un anticuerpo que se une específicamente a un producto génico del gen de CF normal o el gen de CF mutante;
- (b) medios de reactivos para detectar la unión del anticuerpo al producto génico; y
- 30 (c) estando los medios de anticuerpo y reactivos cada uno presentes en cantidades eficaces para llevar a cabo el inmunoensayo.

35 De acuerdo con un aspecto adicional de la invención, se proporciona un estuche para someter a ensayo la presencia de un gen de CF normal correspondiente a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1 o un gen de CF mutante que tiene al menos una deleción de tres pares de bases en el gen de CF normal, deleción que da como resultado la deleción de una fenilalanina procedente del residuo aminoácido en la posición 508 de la figura 1, mediante hibridación, que comprende:

- (a) una sonda de oligonucleótido que se une específicamente al gen de CF normal o al gen de CF mutante;
- 40 (b) medios de reactivos para detectar la hibridación de la sonda de oligonucleótido al gen de CF normal o al gen de CF mutante; y
- (c) estando presentes cada uno de los medios de sonda y reactivos en cantidades eficaces para llevar a cabo el ensayo de hibridación.

45 De acuerdo con un aspecto adicional de la invención, se proporciona un anticuerpo regulador de la conductancia en la membrana anti-fibrosis quística (CFTR), policlonal o monoclonal, inmunológicamente activo, específico para un polipéptido CFTR de acuerdo con la invención. Es un aspecto adicional de la invención un hibridoma que produce el anticuerpo monoclonal de la invención.

50 En otro aspecto, la invención comprende una composición para el tratamiento de la fibrosis quística en un paciente que comprende una cantidad terapéuticamente eficaz de un polipéptido CFTR normal purificado de la invención. El vehículo puede incluir una proteína tensioactiva de los pulmones para facilitar la aplicación de la composición a células epiteliales respiratorias.

55 De acuerdo con otro aspecto de la invención, se proporciona una composición para la terapia génica de la fibrosis quística, que comprende una molécula de ADN de acuerdo con la invención y un vehículo para suministrar dicha molécula de ADN a una célula de un paciente de fibrosis quística. El vehículo puede ser, por ejemplo, un vector recombinante.

60 De acuerdo con otro aspecto de la invención, se proporciona un animal no humano que comprende un sistema de células heterólogo que comprende un vector de clonación recombinante de acuerdo con la invención, vector que induce síntomas de fibrosis quística en dicho animal.

65 De acuerdo con otro aspecto de la invención, un ratón transgénico que exhibe síntomas de fibrosis quística se puede obtener mediante las etapas que comprenden introducir un gen de fibrosis quística que codifica una proteína reguladora de la conductancia en la membrana de la fibrosis quística mutante que incluye al menos una deleción de fenilalanina en la posición correspondiente a 508 en la figura 1 en un oocito de ratón, o un embrión de ratón;

ES 2 210 224 T3

y opcionalmente, inactivar el gen de la fibrosis quística de ratón endógeno.

Con el fin de entender con mayor detalle la invención se hará ahora referencia a los dibujos que se acompañan, en los que:

5 la figura 1 es la secuencia de nucleótidos del gen de CF y la secuencia de aminoácidos de la proteína CFTR;

la figura 2 es un mapa de restricción del gen de CF y la estrategia esquemática utilizada para el desplazamiento y el salto del cromosoma al gen;

10 la figura 3 es un mapa de electroforesis en gel de campo pulsado de la región que incluye y rodea al gen de CF;

las figuras 4A, 4B y 4C muestran la detección de secuencias de nucleótidos conservadas mediante hibridación en especies cruzadas;

15 la figura 4D es un mapa de restricción de segmentos solapantes de sondas E4.3 y H1.6;

la figura 5 es un análisis de hibridación de borrones de ARN utilizando sondas genómicas y de ADNc. Se muestra la hibridación de ARN a fibroblastos, tráquea (normal y CF), páncreas, hígado, HL60, T84 y cerebro;

20 la figura 6 es el estado de metilación de la región clonada E4.3 en el extremo 5' del gen de CF;

la figura 7 es un mapa de restricción del ADNc de CFTR que muestra la alineación del ADNc con respecto a los fragmentos de ADN genómico;

25 la figura 8 es un análisis de los borrones de gel de ARN que describe la hibridación mediante una porción del ADNc de CFTR (clon 10-1) a un transcrito de ARNm de 6,5 kb en diversos tejidos humanos;

30 la figura 9 es un análisis de hibridación de un borrón de ADN que representa la hibridación mediante los clones de ADNc de CFTR al ADN genómico digerido con Eco RI y Hind III;

la figura 10 es un experimento de extensión del cebador que caracteriza los extremos 5' y 3' del ADNc de CFTR;

35 la figura 11 es un perfil de hidropatía y muestra estructuras secundarias predichas de CFTR;

la figura 12 es un análisis mediante matriz de puntos de homologías internas en el polipéptido de CFTR predicho;

la figura 13 es un modelo esquemático de la proteína de CFTR predicha;

40 la figura 14 es un diagrama esquemático de polimorfismos de longitud de fragmentos de restricción (RFLP's, del inglés restriction fragment length polymorphisms) estrechamente ligados al gen de CF, en el que el triángulo invertido indica la localización de la delección de tres pares de bases F508;

45 la figura 15 representa la detección de la mutación de F508 mediante hibridación de oligonucleótidos, detectando la sonda N la secuencia normal y detectando la sonda F la secuencia mutante de CF;

la figura 16 representa la alineación de los segmentos más conservados de los NBFs extendidos de CFTR con regiones equiparables de otras proteínas;

50 la figura 17 es la secuencia de ADN en torno a la delección F508;

la figura 18 es una representación del gel de secuenciación de nucleótidos que muestra la secuencia de ADN en la delección F508;

55 las figuras 19A y 19B son geles de poliacrilamida teñidos de azul Coomassie después de la electroforesis de proteínas procedentes de lisados bacterianos (JM 101), bacterias que fueron transformadas con los plásmidos pGEX;

la figura 20 son inmuno-borrones de lisados bacterianos que contienen la proteína de fusión nº 1 (en la Tabla 8) con sueros pre-inmunes e inmunes procedentes de dos conejos diferentes;

60 La figura 21 es un inmuno-borrón de membranas T-84 utilizando suero inmune procedente del conejo nº 1 de la figura 20; y

65 la figura 22 son borrones de inmuno-puntos sondados con sueros pre-inmunes e inmunes de un conejo inmunizado con el conjugado KLH del péptido nº 2 de la Tabla 8.

Con el fin de poder entender con mayor detalle la invención, se describirán ahora realizaciones de la misma, sólo a modo de ilustración.

ES 2 210 224 T3

1. Definiciones

Con el fin de facilitar la revisión de las diversas realizaciones de la invención y comprender los diversos elementos y constituyentes utilizados en la realización de la invención y utilizando la misma, la siguiente definición de términos y expresiones utilizados en la descripción de la invención es como sigue:

CF - fibrosis quística

Portador de CF - una persona de salud aparente cuyos cromosomas contiene un gen de CF mutante que puede ser transmitido a la descendencia de esa persona.

Paciente de CF - una persona que porta un gen de CF mutante en cada cromosoma, de modo de exhiben los síntomas clínicos de fibrosis quística.

Gen de CF - el gen cuyas formas mutantes se asocian con la enfermedad fibrosis quística. Se entiende que esta definición incluye los diversos polimorfismos de la secuencia que existen, en donde sustituciones de nucleótidos en la secuencia del gen no afectan a la función esencial del producto génico. Este término se refiere principalmente a una secuencia codificada aislada, pero también puede incluir alguna o la totalidad de los elementos y/o intrones reguladores flanqueantes.

CF-PI - fibrosis quística pancreática insuficiente, el subgrupo clínico principal de pacientes de fibrosis quística, caracterizados por una función exocrina pancreática insuficiente.

CF-PS - fibrosis quística pancreática suficiente, un subgrupo clínico de pacientes de fibrosis quística con una función exocrina pancreática suficiente para la digestión normal de los alimentos.

CFTR - proteína reguladora de la conductancia de la membrana de fibrosis quística, codificada por el gen de CF. Esta definición incluye la proteína según se aísla de fuentes humanas o animales, según se produce por organismos recombinantes, y según se sintetiza química o enzimáticamente. Se entiende que esta definición incluye las diversas formas polimorfas de la proteína en donde sustituciones de aminoácidos en las regiones variables de la secuencia no afectan al funcionamiento esencial de la proteína o su perfil hidropático o estructura secundaria o terciaria.

ADN - la nomenclatura convencional se utiliza para identificar las bases.

ADN sin intrones - un trozo de ADN que carece de los segmentos no codificadores internos, por ejemplo ADNc.

Secuencia del locus IRP - (relacionada con proto-oncogen int-1), un gen localizado próximo al gen de CF.

CFTR mutante - una proteína que es muy análoga a CFTR en términos de estructuras primaria, secundaria y terciaria, pero en donde un pequeño número de sustituciones y/o deleciones y/o inserciones de aminoácidos dan como resultado un deterioro de su función esencial, de modo que organismos cuyas células epiteliales expresan CFTR mutante más que CFTR muestran los síntomas de la fibrosis quística.

mCF - un gen de ratón ortólogo al gen de CF humano

NBFs - pliegues de unión de nucleótidos (ATP)

ORF - marco de lectura abierta

PCR - reacción en cadena de la polimerasa

Proteína - se utiliza una nomenclatura convencional de una sola letra para identificar los aminoácidos

Dominio R - un dominio citoplásmico muy cargado de la proteína CFTR

RSV - virus del sarcoma de Rous

SAP - proteína tensioactiva

RFLP - polimorfismo de la longitud del fragmento de restricción

2. Aislamiento del gen de CF

Utilizando el desplazamiento y salto de cromosomas y la hibridación de ADNc, se han aislado secuencias de ADN que abarcan > 500 pares de kilobases (kb) a partir de una región en el brazo largo del cromosoma 7 humano que contiene el gen de la fibrosis quística (CF). En esta región se han identificado varias secuencias transcritas y segmentos conservados. Uno de éstos corresponde al gen de CF y abarca aproximadamente 250 kb de ADN genómico. Clones de ADN complementario (ADNc) solapantes han sido aislados de bancos de células epiteliales con un segmento

de ADN genómico que contiene una porción del gen de la fibrosis quística. La secuencia de nucleótidos del ADNc aislado se muestra en la figura 1. En cada fila de las respectivas secuencias, la fila inferior es una lista, en nomenclatura convencional, de la secuencia de nucleótidos. La fila superior en cada fila respectiva de secuencias es una nomenclatura convencional de una sola letra para el aminoácido correspondiente al respectivo codón.

Transcritos de un tamaño de aproximadamente 6500 nucleótidos se pueden detectar en tejidos afectados en pacientes con CF. En base a la secuencia de nucleótidos aislada, la proteína predicha consiste en dos regiones similares, cada una de las cuales contiene un primer dominio que tiene propiedades consistentes con la asociación de la membrana y un segundo dominio que se piensa está implicado en la unión a ATP.

Una delección de 3 pb da como resultado la omisión de un residuo fenilalanina en el centro del primer dominio de unión de nucleótidos predicho (posición del aminoácido 508 del producto génico CF) ha sido detectada en pacientes de CF. Esta mutación en la secuencia de ADN normal de la figura 1 corresponde a aproximadamente el 70% de las mutaciones en pacientes de fibrosis quística. Datos de haplotipo ampliados, basados en marcadores de ADN estrechamente enlazados al supuesto gen de la enfermedad sugieren que el resto de la agrupación de genes mutantes de CF consiste en múltiples mutaciones diferentes. Un pequeño conjunto de estos últimos alelos mutantes (aproximadamente el 8%) puede conferir una función exocrina pancreática residual en un subgrupo de pacientes que son pancreáticos suficientes.

2.1. Marcha y salto de los cromosomas

Grandes cantidades del ADN que rodea a las regiones de enlace D7S122 y D7S340 de Rommens et al *supra* se investigaron en cuanto a secuencias de genes candidatas. Además de los métodos de marcha de cromosomas convencionales, se emplearon técnicas de salto de cromosomas para acelerar el proceso de investigación. A partir de cada punto extremo de salto se podría iniciar una nueva marcha bidireccional. Marchas secuenciales retenidas por regiones "no clonables", con las que a menudo se topa en el genoma de mamíferos, podrían ser evitadas mediante un salto del cromosoma.

El banco de saltos de cromosomas utilizado ha sido previamente descrito [Collins et al, *Science* 325, 1046 (1987); Ianuzzi et al, *Am. J. Hum. Genet.* 44, 695 (1989)]. El banco original se preparó a partir de un gel de campo pulsado preparativo y se pretendió que contuviera fragmentos parciales de Eco RI de 70-130 kb; la subsiguiente experiencia con este banco indica que también estaban representados fragmentos más pequeños, y se han encontrado tamaños de saltos de 25-110 kb. El banco se extendió en placas sobre MC1061 huésped sup⁻ y se rastreó mediante técnicas convencionales [Maniatis et al]. Los clones positivos se subclonaron en pBRA23Ava, y el comienzo y final del salto se identificó mediante digestión con Eco RI y Ava I según se describe en Collins, *Genome analysis: A practical approach* (IRL, Londres, 1988), págs. 73-94). Para cada clon, se verificó un fragmento desde el extremo del salto para confirmar su localización en el cromosoma 7. La región del cromosoma contigua cubierta por la marcha y salto del cromosoma era de aproximadamente 250 kb. La dirección de los saltos se desvió mediante una elección cuidadosa de las sondas, según se describe por Collins et al e Ianuzzi et al, *supra*. La región completa clonada, incluidas las secuencias aisladas con el uso del ADNc del gen de CF, es de aproximadamente 500 kb.

La representación esquemática de la estrategia de marcha y salto de los cromosomas se ilustra en la figura 2. Los exones del gen de CF se indican mediante números romanos en esta figura. Las líneas horizontales por encima del mapa indican las etapas de marcha, mientras que los arcos por encima del mapa indican las etapas de salto. La figura avanza de izquierda a derecha en cada una de las seis filas, con la dirección de los extremos hacia 7cen y 7qter como se indica. El mapa de restricción para las enzimas Eco RI, Hind III y Bam HI se muestra por encima de la línea continua, que abarca la región completa clonada. Sitios de restricción indicados con flechas más que líneas verticales indican sitios que no han sido inequívocamente posicionados. Sitios de restricción adicionales para otras enzimas se muestran por debajo de la línea. Los huecos en la región clonada se indican mediante.

Estos se producen únicamente en la porción detectada por los clones de ADNc del transcrito de CF. En base a la representación en mapa de campo pulsado de la región, no es probable que los huecos sean grandes. Los clones de marcha, según se indica por las flechas horizontales por encima del mapa, tienen la dirección de la flecha que indican el progreso de la marcha obtenido con cada uno de los clones. Los clones cósmidos comienzan con la letra c; todos los otros clones son fagos. El cósmido CF26 demostró ser una quimera; la porción de línea discontinua se deriva de un fragmento genómico diferente en otro cromosoma. Los números romanos I a XXIV indican la localización de exones del gen de CF. Los recuadros horizontales mostrados por encima de la línea son sondas utilizadas durante los experimentos. Tres de las sondas representan la subclonación independiente de fragmentos previamente identificados para detectar polimorfismos en esta región: H2.3A corresponde a la sonda XV2C (X. Estivill et al, *Nature*, 326: 840 (1987)), la sonda E1 corresponde a KM19 (Estivill, *supra*), y la sonda E4.1 corresponde a Mp6d.9 (X. Estivill et al, *Am. J. Hum. Genet.* 44, 704 (1989)). G-2 es un subfragmento de E6 que detecta una secuencia transcrita. R161, R159 y R160 son oligonucleótidos sintéticos construidos a partir de partes de la secuencia del locus de IRP [B. J. Wainwright et al, *EMBO J.*, 7:1743 (1988)], indicando la localización de este transcrito en el mapa genómico.

Dado que los dos marcadores de ADN independientemente aislados, D7S122 (pH131) y D7S340 (TM58), estaban únicamente separados aproximadamente 10 kb (figura 2), las marchas y los saltos se iniciaron esencialmente desde un único punto. La dirección de la marcha y salto con respecto a MET y D7S8 se estableció luego con el cruzamiento de varios sitios de reconocimiento de la endonucleasa de restricción que corta contadas veces (tales como para Xho

I, Nru I y Not I, véase la figura 2) y con referencia al mapa físico de largo intervalo de J. M. Rommens et al. Am. J. Hum. Genet., en prensa; A. M. Poustka, et al, Genomics 2, 337 (1988); M. L. Drumm et al. Genomics 2, 346 (1988). Los datos de la representación en mapa de campo pulsado revelaron también que el sitio Not I identificado por los autores de la presente invención (véase la figura 2, posición 113 kb) correspondía al previamente encontrado asociado con el locus IRP (Estivill et al 1987, *supra*). Dado que los subsiguientes estudios genéticos demostraron que CF estaba localizado, lo más probablemente, entre IRP y D7S8 [M. Farrall et al, Am. J. Hum. Genet. 43, 471 (1988), B. - S. Kerem et al. Am. J. Hum. Genet. 44, 827 (1989)], el esfuerzo de marcha y salto continuó exclusivamente hacia la clonación de este intervalo. Sin embargo, se aprecia que se localizaron e investigaron ampliamente otras regiones codificantes, según se identifica en la figura 2, por ejemplo G-2, CF14 y CF16. Investigaciones extensas de este tipo de estas otras regiones revelaron que éstas no eran el gen de CF basado en los datos genéticos y en el análisis de la secuencia. Dada la carencia de conocimientos de la localización del gen de CF y sus características, el examen extenso y laborioso de las presuntas regiones codificantes contiguas no avanzó la dirección de investigación del gen de CF. Sin embargo, estas investigaciones eran necesarias con el fin de excluir la posibilidad de que el gen de CF se encontrara en estas regiones.

Se encontró que tres regiones en el segmento de 280 kb no eran fácilmente recuperables en los bancos genómicos amplificados inicialmente utilizados. Estas regiones menos clonables estaban situadas próximas a los segmentos de ADN H2.3A y X.6, y justo más allá del cósmido cW44, en las posiciones 75-100 kb, 205-225 kb y 275-285 kb en la figura 2, respectivamente. Se encontró que los clones recombinantes próximos a H2.3A eran muy inestables con reorganizaciones drásticas sólo después de unos cuantos pasos de cultivo bacteriano. Para rellenar los huecos resultantes, se construyeron bancos de marcha primarios utilizando sistemas de huésped-vector especiales, de los que se ha reseñado permiten la propagación de secuencias inestables [A. R. Wyman, L. B. Wolfe, D. Botstein, Proc. Nat. Acad. Sci. U. S. A. 82, 2880 (1985); K. F. Wertman, A. R. Wyman, D. Botstein, Gene 49, 253 (1986); A. R. Wyman, K. F. Wertman, D. Barker, C. Helms, W. H. Petri, Gene, 49, 263 (1986)]. A pesar de que la región próxima al cósmido cW44 permanece sin recuperar, la región próxima a X.6 se rescató con éxito con estos bancos.

2.2 Construcción de bancos genómicos

Los bancos genómicos se construyeron según los procesos descritos en Maniatis, et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York 1982) y se listan en la Tabla I. Esta incluye bancos de ocho fagos, uno de los cuales fue proporcionado por T. Maniatis [Fritsch et al, Cell, 19:959 (1989)]; el resto se construyó como parte de este trabajo de acuerdo con los procesos descritos en Maniatis et al, *supra*. Se clonaron cuatro bancos de fagos en λ DASH (comercialmente disponible de Stratagene) y tres en λ FIX (comercialmente disponible de Stratagene), con los brazos del vector proporcionados por el fabricante. Se construyó un banco λ DASH a partir de ADN parcialmente digerido con Sau 3A procedente de un híbrido humano-hámster que contenía el cromosoma 7 humano (4AF/102/K015) [Rommens et al Am. J. Hum. Genet 43, 4 (1988)] y otros bancos a partir de la digestión parcial con Sau 3A, total con Bam HI o total con Eco RI de la sangre periférica humana o de ADN linfoblastoide. Para evitar la pérdida de secuencias inestables, se propagaron cinco de los bancos de fagos en los huéspedes DB1316 (recD⁻), CES 200 (recBC⁻) [Wyman et al, *supra*, Wertman et al *supra*, Wyman et al *supra*]; o TAP90 [Patterson et al Nucleic Acids Res. 15:6298 (1987)] deficientes en la recombinación. Luego se construyeron tres bancos de cósmidos. En uno de ellos, se utilizó el vector pCV108 [Lau et al Proc. Natl. Acad. Sci USA 80:5225 (1983)] para clonar ADN parcialmente digerido (Sau 3A) a partir de 4AF/102/K015 [Rommens et al Am. J. Hum. Genet. 43:4 (1988)]. Se preparó un segundo banco de cósmidos clonando ADN linfoblastoide humano parcialmente digerido (Mbo I) en el vector pWE-IL2R, preparado al insertar el ADNc impulsado por el promotor de RSV (virus del sarcoma de Rous) para la cadena α del receptor de interleuquina-2 (suministrada por M. Fordis y B. Howard) en el lugar del gen de resistencia a *neo* de pWE15 [Wahl et al Proc. Natl. Acad. Sci. USA 84:2160 (1987)]. Se preparó un banco de cósmidos adicional parcial de Mbo I en el vector pWE-IL2-Sal, creado al insertar un enlazador Sal I en el sitio de clonación Bam HI de pWE-EL2R (M. Drumm, datos no publicados); esto permite el uso de la técnica de relleno parcial para ligar los extremos Sal I y Mbo I, evitando inserciones en tandem [Zabarovsky et al Gene 42:19 (1986)]. Los bancos de cósmidos se propagaron en cepas huésped de *E. coli* DH1 o 490A [M. Steinmetz, A. Winoto, K. Minard, L. Hood, Cell 28, 489 (1982)].

ES 2 210 224 T3

TABLA 1

Bancos genómicos

5	<u>Vector</u> <u>Ref.</u>	<u>Fuente de ADN humano</u>	<u>Huésped</u>	<u>Complejidad</u>
	λ Charon 4A	ADN total de hígado humano parcialmente digerido con Hae II/Alu I	LE392	1 x 10 ⁶ Lawn et al 1980
10				
	pCV108	ADN parcialmente digerido con Sau 3A procedente de 4AF/KO15	DK1	3 x 10 ⁶ (amplificado)
15				
	λdash	ADN parcialmente digerido con Sau 3A procedente de 4AF/KO15	LE392	1 x 10 ⁶ (amplificado)
20				
	λdash	ADN total de la sangre periférica humana parcialmente digerido con Sau 3A	DB1316	1,5 x 10 ⁶
25				
	λdash	ADN total de la sangre periférica humana parcialmente digerido con Bam HI	DB1316	1,5 x 10 ⁶
30				
	λdash	ADN total de la sangre periférica humana parcialmente digerido con Eco RI	DB1316	8 x 10 ⁶
35				
	λFIX	ADN linfoblastoide humano parcialmente digerido con Mbo I	LE392	1,5 x 10 ⁶
40				
	λFIX	ADN linfoblastoide humano parcialmente digerido con Mbo I	CE200	1,2 x 10 ⁶
45				
	λFIX	ADN linfoblastoide humano parcialmente digerido con Mbo I	TAP90	1,3 x 10 ⁶
50				
	pWE-IL2R	ADN linfoblastoide humano parcialmente digerido con Mbo I	490A	5 x 10 ⁶
55				
	pWE-IL2R-Sal	ADN linfoblastoide humano parcialmente digerido con Mbo I	490A	1,2 x 10 ⁶
60				
	λCh3A Δlac (salto)	ADN linfoblastoide humano parcialmente digerido con Mbo I (24-110 kb)	MC1061	3 x 10 ⁶ Collins et al supra e Ianuzzi et al supra
65				

ES 2 210 224 T3

Tres de los bancos de fagos se propagaron y amplificaron en la cepa bacteriana LE392 de *E. coli*. Cuatro bancos subsiguientes se extendieron en placas sobre los huéspedes DB1316 (recD⁻) o CES200 (rec BC⁻) deficientes en la recombinación [Wyman 1985, *supra*; Wertman 1986, *supra*; y Wyman 1986, *supra*] o en un caso TAP90 [T.A. Patterson y M. Dean, *Nucleic Acids Research* 15, 6298 (1987)].

Se purificaron segmentos de ADN de una copia sencilla (exentos de elementos repetitivos) próximos a los extremos de cada inserto de fago o cósmido y se utilizaron como sondas del rastreo de bancos para aislar fragmentos de ADN solapantes mediante procesos convencionales. (Maniatis, et al, *supra*).

1-2 x 10⁶ clones de fagos se extendieron en 25-30 placas de Petri de 150 mm con el huésped bacteriano indicador apropiado y se incubaron a 37°C durante 10-16 h. Se prepararon “elevadores” por duplicado para cada placa con membranas de nitrocelulosa o nilón, se prehibridaron e hibridaron en las condiciones descritas [Rommens et al, 1988, *supra*]. Las sondas se marcaron con ³²P hasta una actividad específica de > 5 x 10⁸ cpm/μg utilizando el proceso de imprimación aleatorio [A. P. Feinberg y B. Vogelstein, *Anal. Biochem.* 132, 6 (1983)]. El banco de cósmidos se esparció sobre placas que contenían ampicilina y se rastreo de una manera similar.

Las sondas de ADN que dieron altas señales de fondo pudieron utilizarse de nuevo con más éxito pre-asociando la sonda en ebullición con 250 μg/ml de ADN de la placenta desnaturalizada, cizallada durante 60 minutos antes de añadir la sonda a la bolsa de hibridación.

Para cada etapa de marcha, la identidad del fragmento de ADN clonado se determinó mediante hibridación con un panel híbrido de células somáticas para confirmar su localización en el cromosoma, y mediante representación en mapa de restricción y análisis de transferencia Southern para confirmar su colinearidad con el genoma.

La región clonada combinada total de las secuencias de ADN genómico aisladas y los clones de ADNc solapantes se extendieron > 500 kb. Para asegurar que los segmentos de ADN aislados por los procesos de marcha y salto del cromosoma fueran colineares con la secuencia genómica, cada segmento se examinó:

- (a) mediante análisis de hibridación con líneas de células híbridas somáticas de seres humanos-roedores para confirmar la localización del cromosoma 7,
- (b) electroforesis en gel de campo pulsado, y
- (c) comparación del mapa de restricción del ADN clonado con el del ADN genómico.

Por consiguiente, secuencias de ADN humano de una copia sencilla se aislaron de cada fago recombinante y del clon del cósmido y se utilizaron como sondas en cada uno de estos análisis de hibridación según se realiza por el proceso de Maniatis, et al, *supra*.

Mientras que la mayoría de los materiales aislados de fagos y cósmidos representaban clones de marcha y salto correctos, unos pocos procedían de artefactos de clonación o secuencias de hibridación cruzada procedentes de otras regiones en el genoma humano o del genoma de hámster en los casos en los que los bancos se derivaban de una línea de células híbridas de ser humano-hámster. La confirmación de la localización correcta era particularmente importante para clones aislados mediante el salto del cromosoma. Se consideraron muchos clones de salto y dieron como resultado una información no conclusiva que conducía a la dirección de investigación fuera del gen.

2.3 Confirmación del mapa de restricción

La confirmación adicional del mapa físico global de los clones solapantes se obtuvo mediante un análisis de representación en mapa de restricción de intervalo amplio con el uso de electroforesis en gel de campo pulsado (J. M. Rommens, et al. *Am. J. Hum. Genet.* en prensa, A. M. Poustka et al, 1988, *supra*, M. L. Drumm et al, 1988 *supra*).

Las figuras 3A a 3E ilustran los hallazgos del estudio de la representación en mapa de restricción de amplio intervalo, en el que en el panel E se da una representación esquemática de la región. El ADN procedente de la línea de células de ser humano-hámster 4AF/102/K015 se digirió con las enzimas (A) Sal I, (B) Xho I, (C) Sfi I y (D) Nae I, se separó mediante electroforesis en gel de campo pulsado y se transfirió a Zetaprobe[®] (BioRad). Para cada enzima, un borrón sencillo se hibridó secuencialmente con las sondas indicadas por debajo de cada uno de los paneles de las figuras A a D, con una separación por arrastre del borrón entre hibridaciones. Los símbolos para cada enzima de la figura 3E son: A, Nae I; B, Bss HII; F, Sfi I; L, Sal I; M, Mlu I; N, Not I; R, Nru I; y X, Xho I. C corresponde a la región de la zona de compresión del gel. Los preparados de ADN, la digestión por restricción y los métodos de electroforesis en gel de campo cruzado han sido descritos (Rommens et al, en prensa, *supra*). Los geles en la figura 3 se extendieron en 0,5 x TBE a razón de 7 voltios/cm durante 20 horas con una conmutación de una rampa lineal de 10-40 segundos para (A), (B) y (C), y a razón de 8 voltios/cm durante 20 horas con una conmutación con una rampa lineal de 50-150 segundos para (D). Para cada panel se indican a continuación interpretaciones esquemáticas del modelo de hibridación. Los tramos de fragmento se dan en kilobases y se dimensionaron mediante comparación con el bacteriófago λADN oligomerizado y cromosomas de *Saccharomyces cerevisiae*.

H4.0, J44, EG1.4 son sondas genómicas generadas a partir de los experimentos de marcha y salto (véase la figura

2). J30 ha sido aislado mediante cuatro saltos consecutivos desde D7S8 (Collins et al, 1987, *supra*; Ianuzzi et al, 1989, *supra*; M. Dean et al, sometido a publicación). 10-1, B.75 y CE1.5/1.0 son sondas de ADNc que cubren diferentes regiones del transcrito de CF: 10-1 contiene los exones I-VI, B.75 contiene los exones V-XII y CE1.5/1.0 contiene los exones XII-XXIV. En la figura 3E se muestra un mapa compuesto del intervalo completo de MET - D7S8. La región
 5 encuadrada indica el segmento clonado mediante marcha y salto, y la porción de barras indica la región cubierta por el transcrito de CF. La región rica en CpG asociada con el locus D7S23 (Estivill et al, 1987, *supra*) se encuentra en el sitio Not I mostrado entre paréntesis. Éste y otros sitios mostrados entre paréntesis o corchetes no se cortan en 4AF/102/K015, sino que se han observado en líneas de células linfoblastoides humanas.

10 2.4 Identificación del gen de CF

Basado en los hallazgos de la representación en mapa de restricción de amplio intervalo detallada anteriormente, se determinó que el gen de CF completo está contenido en un fragmento Sal I de 380 kb. La alineación de los sitios de restricción derivados del análisis en gel de campo pulsado con los identificados en los clones de ADN genómico
 15 parcialmente solapantes revelaron que el tamaño del gen de CF era de aproximadamente 250 kb.

La enzima de restricción más informativa que servía para alinear el mapa de los fragmentos de ADN clonados y el mapa de restricción de amplio intervalo era Xho I; la totalidad de los 9 sitios Xho I identificados con los clones de ADN recombinante parecieron ser susceptibles a al menos una escisión parcial en el ADN genómico (compárense
 20 los mapas de las figuras 1 y 2). Además, análisis de hibridación con sondas derivadas del extremo 3' del gen de CF identificaron 2 sitios SfiI y confirmaron la posición de un sitio Nae I anticipado.

Estos hallazgos sustentan además la conclusión de que los segmentos de ADN aislados por procesos de marcha y salto de cromosoma eran colineares con la secuencia genuina.

25 2.5 Criterios de identificación

Un resultado positivo basado en uno o más de los siguientes criterios sugirió que un segmento de ADN clonado puede contener secuencias de genes candidatos:

- 30 (a) detección de secuencias de hibridación cruzada en otras especies (dado que muchos genes muestran una conservación evolutiva),
- 35 (b) identificación de islas de CpG que a menudo marcan el extremo 5' de los genes de vertebrados [A. P. Bird, *Nature*, 321, 209 (1986); M. Gardiner-Garden y M. Frommer, *J. Mol. Biol.* 196, 261 (1987)],
- (c) examen de posibles transcritos de ARNm en tejidos afectados en pacientes de CF,
- 40 (d) aislamiento de secuencias de ADNc correspondientes,
- (e) identificación de marcos de lectura abierta mediante secuenciación directa de segmentos de ADN clonados.

La hibridación de especies cruzadas mostró una fuerte conservación de la secuencia entre ADN humano y bovino cuando como sondas se utilizaron CF14, E4.3 y H1.6, cuyos resultados se muestran en las figuras 4A, 4B y 4C.

ADNs genómicos humanos, bovinos, de ratón, hámster y pollo se digirieron con Eco RI (R), Hind III (H) y Pst I (P), se sometieron a electroforesis y se transfirieron a Zetabind® (BioRad). Los procesos de hibridación de Rommens et al, 1988, *supra*, se utilizaron con el lavado más estricto a 55°C, 0,2 X SSC y SDS al 0,1%. Las sondas utilizadas para la hibridación, en la figura 4, incluyen: (A) el cósmido CF14 completo, (B) E4.3, (C) H1.6. En la vista esquemática de
 50 la figura (D) la región sombreada indica la zona de la conservación de especies cruzadas.

El hecho de que en ADN bovino se detectaran diferentes conjuntos de bandas con dos segmentos de ADN solapantes (H1.6 y E4.3) sugirió que las secuencias conservadas estaban localizadas en los límites de la región solapada (figura 4, (D)). Cuando estos segmentos de ADN se utilizaron para detectar transcritos de ARN procedentes de una diversidad de tejidos, no se detectó ninguna señal de hibridación. En un intento de comprender la región de hibridación cruzada y de identificar posibles marcos de lectura abierta, se determinaron las secuencias de ADN del H1.6 completo y parte del fragmento E4.3. Los resultados mostraron que, excepto para un largo tramo de secuencia rica en CG que contenía los sitios de reconocimiento para dos enzimas de restricción (Bss HII y Sac II), que se encuentran a menudo asociadas con islas CpG metiladas, sólo existían cortos marcos de lectura abierta que no podían explicar con facilidad
 60 las fuertes señales de hibridación de especies cruzadas.

Para examinar el estado de metilación de esta región muy rica en CG revelada mediante secuenciación, muestras de ADN genómico preparadas a partir de fibroblastos y linfoblastos se digirieron con las enzimas de restricción Hpa II y Msp I y se analizaron mediante hibridación de borrón en gel. La enzima Hpa II corta la secuencia de ADN 5'-CCGG-3' sólo cuando la segunda citosina no está metilada, mientras que Msp I corta esta secuencia independientemente del estado de metilación. Para ambas enzimas se generaron pequeños fragmentos de ADN, indicando que esta región rica en CpG está de hecho submetilada en ADN genómico. La hibridación por borrón en gel con el segmento E4.3

ES 2 210 224 T3

(figura 6) revela fragmentos de hibridación pequeños con ambas enzimas, indicando la presencia de una isla CpG hipometilada.

Los resultados anteriores sugieren fuertemente la presencia de una región codificadora en este locus. Dos segmentos de ADN (E4.3 y H1.6) que detectaban señales de hibridación de especies cruzadas procedentes de esta zona se utilizaron como sondas para rastrear bancos de ADNc preparados a partir de varios tipos de tejidos y células.

Los bancos de ADNc procedentes de células epiteliales cultivadas se prepararon como sigue. Células de las glándulas sudoríparas derivadas de un individuo no CF y de un paciente CF se desarrollaron hasta un primer paso según se describe [G. Collie et al, *In Vitro Cell. Dev. Biol.* 21, 592, 1985]. En estas células se confirmó la presencia de canales rectificadores hacia afuera (J. A. Tabcharani, T. J. Jensen, J. R. Riordan, J. W. Hanrahan, *J. Memb. Biol.*, en prensa), pero las células CF eran sensibles a la activación por parte de AMP cíclico (T. J. Jensen, J. W. Hanrahan, J.A. Tabcharani, M. Buchwald y J. R. Riordan, *Pediatric Pulmonology*, Suplemento 2, 100, 1988). El ARN se aisló de ellas mediante el método de J. M. Chirgwin et al (*Biochemistry* 18, 5294, 1979). Se seleccionó ARN de poli A+ (H. Aviv y P. Leder, *Proc. Natl. Acad. Sci. USA* 69, 1408, 1972) y se utilizó como plantilla para la síntesis de ADNc con oligo (dT) 12-18 como cebador. La segunda hebra se sintetizó de acuerdo con Gubler y Hoffman (*Gene* 25, 263, 1983). Ésta se metiló con Eco RI metilasa y los extremos se nivelaron con T4 ADN polimerasa. Los enlazadores Eco RI fosforilados se ligaron al ADNc y se restringieron con Eco RI. La separación de enlazadores en exceso y el fraccionamiento del tamaño parcial se consiguió mediante cromatografía en Biogel A-50. A continuación, los ADNc se ligaron en el sitio Eco RI del lambda ZAP comercialmente disponible. Recombinantes se empaquetaron y propagaron en *E. coli* BB4. Partes de las mezclas de envasado se amplificaron y el resto se retuvo para el rastreo antes de la amplificación. Se utilizaron los mismos procesos para construir un banco a partir de ARN aislado de cultivos pre-confluentes de la línea de células de carcinoma de colon T-84 (Dharmasathaphorn, K. et al. *Am. J. Physiol.* 246, G204, 1984). Los números de recombinantes independientes en los tres bancos eran: 2×10^6 para las células de glándulas sudoríparas no CF, $4,5 \times 10^6$ para las células de las glándulas sudoríparas CF y $3,2 \times 10^6$ de células T-84. Estos fagos se extendieron en placas a razón de 50.000 por cada placa de 15 cm y se realizaron elevaciones de placas utilizando membranas de nilón (Biodyne) y se sondearon con fragmentos de ADN marcados con ^{32}P utilizando ADN polimerasa I y una mezcla aleatoria de oligonucleótidos como cebador. Las condiciones de hibridación estaban de acuerdo con G. M. Wahl y S. L. Berger (*Meth. Enzymol.* 152, 415, 1987). Plásmidos Bluescript[®] se rescataron de clones purificados en placa mediante escisión con fago M13 helper. Los bancos de pulmones y páncreas se adquirieron de Clontech Lab Inc. con tamaños reseñados de $1,4 \times 10^6$ y $1,7 \times 10^6$ clones independientes.

Después de rastrear 7 bancos diferentes cada uno con 1×10^5 - 5×10^6 clones independientes, se aisló un clon único (identificado como 10-1) con H1.6 a partir de un banco de ADNc preparado a partir de las células epiteliales de las glándulas sudoríparas cultivadas de un individuo no afectado (no CF).

El análisis de la secuenciación de ADN demostró que 10-1 contenía un inserto de 920 pb de tamaño y un marco de lectura abierta (ORF, del inglés open reading frame) largo potencial. Dado que un extremo de la secuencia compartía una identidad de secuencia perfecta con H1.6, se concluyó que el clon de ADNc se derivaba probablemente de esta región. Sin embargo, la secuencia de ADN común era sólo de 113 pb de longitud (véanse las figuras 1 y 7). Como se detalla más abajo, esta secuencia corresponde de hecho al exón 5'-próximo del gen de CF putativo. Así, el solapamiento de la secuencia corta explicaba las débiles señales de hibridación en el rastreo del banco y la incapacidad de detectar transcritos en el análisis por borrón en gel de ARN. Además, la orientación de la unidad de transcripción se estableció provisionalmente sobre la base de la alineación de la secuencia de ADN genómico con el presunto ORF de 10-1.

Dado que se estimó que el correspondiente transcrito tenía una longitud de aproximadamente 6500 nucleótidos mediante experimentos de hibridación por borrón en gel de ARN, se requirió un rastreo del banco de ADNc adicional con el fin de clonar el resto de la región codificante. Como resultado de varios rastreos sucesivos con bancos de ADNc generados a partir de la línea de células de carcinoma de colon T84, células de las glándulas sudoríparas normal y CF, páncreas y pulmones de adultos, se aislaron 18 clones adicionales (figura 7, como subsiguientemente se comenta con mayor detalle). El análisis de la secuencia de ADN reveló que ninguno de estos clones de ADNc correspondía a la longitud del transcrito observado, pero fue posible derivar una secuencia consensuada basada en las regiones de solapamiento. Clones de ADN adicionales correspondientes a los extremos 5' y 3' del transcrito se derivaron de los experimentos de la extensión del cebador en 5' y 3'. Juntos, estos clones abarcan un total de aproximadamente 6,1 kb y contienen un ORF capaz de codificar un polipéptido de 1480 residuos aminoácidos (figura 1).

Era inusual observar que la mayoría de los clones de ADNc aislados en este caso contenían inserciones en la secuencia en diversas localizaciones del mapa de restricción de la figura 7. El mapa detalla la estructura genómica del gen de CF. Se indican límites de exón/intrón en los que la totalidad de los clones de ADNc aislados se representan esquemáticamente en la mitad superior de la figura. Muchas de estas secuencias extras correspondían claramente a regiones de intrón transcritas inversamente durante la construcción del ADNc, según se reveló tras la alineación con secuencias de ADN genómico.

Dado que el número de clones de ADNc recombinantes para el gen de CF detectado en el rastreo del banco era mucho menor que el que cabría esperar de la abundancia de transcrito estimado a partir de los experimentos de hibridación con ARN, pareció probable que los clones que contuvieran estructuras aberrantes serían preferentemente retenidos, mientras que los clones adecuados se perderían durante la propagación. Consistente con esta interpretación, se observó

ES 2 210 224 T3

un desarrollo deficiente para la mayoría de los clones recombinantes aislados en este estudio, independientemente del vector utilizado.

5 Los procesos utilizados para obtener los extremos 5' y 3' del ADNc eran similares a los descritos (M. Frohman et al, Proc. Nat. Acad. Sci, USA, 85, 8998-9002, 1988). Para los clones del extremo 5', muestras de páncreas total y de ARN de poli A + de T84 se transcribieron inversamente utilizando un cebador (10b) que es específico para el exón 2 de modo similar a como se ha descrito para la reacción de extensión del cebador, excepto que en la reacción se incluyó el trazador radioactivo. Las fracciones recogidas a partir de una columna de perlas de agarosa de la primera síntesis de la hebra se analizaron mediante reacción en cadena de la polimerasa (PCR-del inglés polymerase chain reaction) de fracciones eluidas. Los oligonucleótidos utilizados se encontraban dentro de la secuencia de 10-1 (separados 145 nucleótidos), justo en 5' del cebador de extensión. Las fracciones más tempranas que proporcionaban producto de PCR se agruparon y se concentraron mediante evaporación y subsiguientemente se agregaron con desoxinucleotidil transferasa terminal (BRL Labs.) y dATP según se recomienda por el suministrador (BRL Labs.). A continuación, se llevó a cabo una segunda síntesis de la hebra con Taq polimerasa (Cetus, AmpliTaq®) utilizando un oligonucleótido que contenía una secuencia de enlazador agregada 5'CGGAATTCTCGAGATC(T)₁₂3'.

La amplificación mediante un experimento anclado (PCR) utilizando la secuencia de enlazador y un cebador interno al cebador de extensión que poseía el sitio de restricción Eco RI en su extremo 5'. Después de la restricción con las enzimas Eco RI y Bgl II y de la purificación en gel de agarosa, productos seleccionados en tamaño se clonaron en el plásmido Bluescript KS disponible de Stratagene mediante procesos convencionales (Maniatis et al, *supra*). Esencialmente, la totalidad de los clones recuperados contenían insertos de menos de 350 nucleótidos. Para obtener los clones del extremo 3', se preparó ADNc de primera hebra con la transcripción inversa de 2 µg de ARN de poli A de T84 utilizando el oligonucleótido de enlazador agregado previamente descrito con condiciones similares a las de la extensión del cebador. Después se llevó a cabo la amplificación mediante PCR con el oligonucleótido enlazador y tres oligonucleótidos diferentes correspondientes a secuencias conocidas del clon T16-4.5. Se llevó a cabo una reacción a escala preparativa (2 x 100 µl) con uno de estos oligonucleótidos con la secuencia 5' ATGAAGTCCAAGGATTTAG3'.

Este oligonucleótido se encuentra aproximadamente 70 oligonucleótidos más arriba de un sitio Hind III dentro de la secuencia conocida de T16-4.5. La restricción del producto de PCR con Hind III y Xho I se siguió mediante purificación en gel de agarosa para seleccionar el tamaño de una banda a 1,0 - 1,4 kb. A continuación, el producto se clonó en el plásmido Bluescript KS disponible de Stratagene. Aproximadamente el 20% de los clones obtenidos se hibridaban a la porción del extremo 3' de T16-4.5. 10/10 de los plásmidos aislados a partir de estos clones tenían mapas de restricción idénticos con tamaños del inserto de aproximadamente 1,2 kb. Todas las reacciones de PCR se llevaron a cabo durante 30 ciclos en tampón sugerido por un suministrador de enzimas.

Un cebador de extensión situado 157 nucleótidos (nt) del extremo 5' del clon 10-1 se utilizó para identificar el producto de partida del putativo transcrito de CF. El cebador estaba marcado en su extremo con γ[³²P]ATP a 5000 Curie/mmol y T4 polinucleótido quinasa y se purificó mediante filtración en gel en columna centrifugada. El cebador radiomarcado se reasoció luego con 4-5 µg de ARN de poli A+ preparado a partir de células de carcinoma de colon T-84 en tampón transcriptasa inversa 2X durante 2 h a 60°C. Después de la dilución y adición de la transcriptasa inversa AMV (Life Sciences, Inc.), la incubación prosiguió a 41°C durante 1 hora. Luego, la muestra se ajustó a NaOH 0,4 M y EDTA 20 mM y, finalmente, se neutralizó con NH₄OAc, pH 4,6, fenol, se extrajo con fenol, se precipitó en etanol, se volvió a disolver en tampón con formamida y se analizó en un gel de secuenciación de poli(acrilamida). Detalles de estos métodos han sido descritos (Meth. Enzymol. 152, 1987, comp. S.L. Berger, A.R. Kimmel, Academic Press, N.Y.).

En el panel A de la figura 10 se muestran los resultados del experimento de la extensión del cebador utilizando un cebador de oligonucleótidos de extensión que parte 157 nucleótidos a partir del extremo 5' de 10-1. Como marcador del tamaño se utiliza el bacteriófago ΦX174 marcado en el extremo y digerido con Hae III (BRL Labs). Dos productos principales se observan a los 216 y 100 nucleótidos. La secuencia correspondiente a los 100 nucleótidos en 10-1 corresponde a una secuencia muy rica en GC (11/12), sugiriendo que esto podría ser un sitio de reposo de la transcriptasa inversa. En el panel B de la figura 10 se muestran los resultados de la PCR anclada a 5'. El gel de agarosa al 1,4% mostrado a la izquierda se emborronó y transfirió a una membrana Zetaprobe® (Bio-Rad Lab). La hibridación del borrón de gel de ADN con 10-1 radiomarcado se muestra a la derecha. Se observa que los productos de extensión 5' varían en tamaño desde 170-280 nt con el producto principal en aproximadamente 200 nucleótidos. La pista de control de PCR muestra un fragmento de 145 nucleótidos. Se obtuvo utilizando los oligómeros de ensayo dentro de la secuencia de 10-1. Los marcadores del tamaño mostrados corresponden a tamaños de 154, 220/210, 298, 344, 394 nucleótidos (escalera de 1 kb adquirida de BRL Lab).

La representación esquemática mostrada por debajo del panel B de la figura 10 esboza el proceso para obtener ADNc de doble hebra utilizado para la amplificación y clonación para generar los clones PA3-5 y TB2-7 mostrados en la figura 7. Los experimentos de PCR anclada para caracterizar el extremo 3' se muestran en el panel C. Como se describe en la representación esquemática por debajo de la figura 10C, se utilizaron tres cebadores, cuya posición relativa uno con otro era conocida para la amplificación con ARN de T84 inversamente transcrito según se describe. Estos productos se separaron sobre un gel de agarosa al 1% y se emborronaron sobre una membrana de nilón según se describe antes. La hibridación del borrón de ADN con la porción 3' del clon T16-4.5 proporcionó bandas de tamaños que correspondían a la distancia entre el oligómero específico utilizado y el extremo 3' del transcrito. Estas bandas en las pistas 1, 2a y 3 se muestran esquemáticamente por debajo del panel C en la figura 10. La banda en la pista 3

ES 2 210 224 T3

es débil, ya que solamente 60 nucleótidos de este segmento se solapan con la sonda utilizada. También se indica en la representación esquemática y como se muestra en la pista 2b, el producto generado por restricción del producto de PCR anclada facilita la clonación para generar el clon THZ-4 mostrado en la figura 7.

5 El análisis de hibridación del borrón de ADN genómico digerido con enzimas Eco RI y Hind III sondadas con porciones de ADNc que abarcan el transcrito completo sugieren que el gen contiene al menos 24 exones numerados con los números romanos I a XXIV (véase la figura 9). Estos corresponden a los números 1 a 24 mostrados en la figura 7. El tamaño de la banda se indica en kb.

10 En la figura 7, los recuadros en blanco indican posiciones aproximadas de los 24 exones que han sido identificados mediante el aislamiento de > 22 clones a partir del rastreo de bancos de ADNc y a partir de experimentos de PCR anclada diseñados para clonar los extremos 5' y 3'. También se indican las longitudes en kb de los fragmentos genómicos Eco RI detectados por cada exón. Los recuadros rayados en la figura 7 indican la presencia de secuencias de intrones, y los recuadros punteados indican otras secuencias. En el extremo inferior de la izquierda se representa, mediante el recuadro en negro, la posición relativa del clon H1.6 utilizado para detectar el primer clon 10-1 de ADNc de entre 10⁶
15 fagos del banco de glándulas sudoríparas normales. Como se muestra en las figuras 4(D) y 7, el clon genómico H1.6 se solapa parcialmente con un fragmento Eco RI de 4,3 kb. La totalidad de los clones de ADNc mostrados se hibridaron a ADNc genómico y/o se representaron en un mapa de restricción fino. Se indican ejemplos de los sitios de restricción que se producen dentro de los ADNc y en los fragmentos genómicos correspondientes.

20 Con referencia a la figura 9, el análisis de hibridación incluye sondas; es decir los clones 10-1 de ADNc para el panel A, T16-1 (porción 3') para el panel B, T16-4.5 (porción central) para el panel C y T16-4.5 (porción extrema 3') para el panel D. En el panel A de la figura 9, la sonda de ADNc 10-1 detecta las bandas genómicas para los exones I a VI. La porción 3' de T16-1 generada por la restricción de Nru I detecta los exones IV a XIII como se muestra en el panel B. Esta sonda se solapa parcialmente con 10-1. Los paneles C y D, respectivamente, muestran bandas genómicas detectadas por los fragmentos Eco RI central y extremo 3' del clon T16-4.5. Dos sitios Eco RI se producen dentro de la secuencia de ADNc y escinden los exones XIII y XIX. Como se indica por los exones entre paréntesis, dos bandas Eco RI genómicas corresponden a cada uno de estos exones. Se observó una hibridación cruzada con otros fragmentos genómicos. Estas bandas, indicadas por N, no tienen un origen de cromosoma 7, ya que no aparecen en híbridos de seres humanos-hámster que contienen el cromosoma 7 humano. Se piensa que la banda tenue en el panel D indicada por XI entre corchetes es provocada por la hibridación cruzada de secuencias debida a la homología interna con el ADNc.

35 Dado que 10-1 detectó una fuerte banda tras la hibridación de ARN por borrón de gel a partir de la línea de células de carcinoma de colon T-84, este ADNc se utilizó para rastrear el banco construido a partir de esa fuente. Se obtuvieron quince positivos de los cuales se purificaron y secuenciaron los clones T6, T6/20, T11, T16-1 y T13-1. El rastreo renovado del mismo banco con un fragmento Bam HI-Eco RI de 0,75 kb a partir del extremo 3' de T16-1 proporcionó T16-4.5. Un fragmento Eco RI de 1,8 kb desde el extremo 3 de T16-4.5 proporcionó T8-B3 y T12a, este último contenía una señal de poliadenilación y cola. Simultáneamente, se rastreó un banco de ADNc de pulmón humano; se aislaron muchos clones incluidos los que se muestran aquí con el prefijo "CDL". También se rastreó un banco de páncreas, proporcionando el clon CDPJ5.

40 Para obtener copias de este transcrito a partir de un paciente de CF, un banco de ADNc procedente de ARN de células epiteliales de las glándulas sudoríparas procedentes de un paciente se rastreó con el fragmento Bam HI - Eco RI de 0,75 kb desde el extremo 3' de T16-1 y se aislaron clones C16-1 y C1-1/5 que cubrían todo, a excepción del exón 1. Estos dos clones exhiben ambos una delección de 3 pb en el exón 10 que no está presente en ningún otro clon que contiene ese exón. Varios clones, incluido CDLS26-1 procedente del banco de pulmones y T6/20 y T13-1 aislados de T84 se derivaron de transcritos parcialmente tratados. Esto se confirmó mediante hibridación genómica y mediante secuenciación a través de los límites del exón-intrón para cada clon. T11 contenía también una secuencia adicional en cada extremo. T16-4.5 contenía una pequeña inserción próxima al límite entre los exones 10 y 11 que no correspondían a la secuencia del intrón. Los clones CDLS16A, 11a y 13a procedentes del banco de pulmones contenían también secuencias extrañas de origen desconocido. El clon C16-1 contenía también una corta inserción correspondiente a una porción del transposón γ de *E. coli*. Este elemento no se detectó en los otros clones. Los clones PA3-5 en 5', generados a partir de ARN del páncreas y TB2-7 generados a partir de ARN de T84 utilizando la técnica de PCR anclada tienen secuencias idénticas, excepto por la diferencia en un solo nucleótido en el extremo 5' según se muestra en la figura 1. El clon en 3', THZ-4 obtenido de ARN de T84, contiene la secuencia 3' del transcrito en concordancia con la secuencia genómica de esta región.

55 A partir de clones de ADNc solapantes se generó una secuencia combinada que representa la presunta región codificante del gen de CF. Dado que la mayoría de los clones de ADNc se derivaban aparentemente de transcritos sin procesar, se llevaron a cabo estudios adicionales para asegurar la autenticidad de la secuencia combinada. Primeramente, cada clon de ADNc se sometió a ensayo en cuanto a la localización al cromosoma 7 mediante análisis de hibridación con un híbrido de células somáticas de seres humanos-hámster que contenía un único cromosoma 7 humano y mediante electroforesis en gel de campo pulsado. Para cada clon se realizó también una representación fina en mapa de la enzima de restricción. Mientras que las regiones solapantes eran claramente identificables para la mayoría de los clones, muchos contenían regiones de modelos de restricción únicos.

65 Para caracterizar adicionalmente estos clones de ADNc, éstos se utilizaron como sondas en experimentos de hibri-

dación en gel con ADN genómico humano digerido con Eco RI o Hind III. Como se muestra en la figura 9, se pudieron detectar cinco a seis fragmentos de restricción diferentes con el ADNc de 10-1 y un número similar de fragmentos con otros clones de ADNc, sugiriendo la presencia de múltiples exones para el gen de CF putativo. Los estudios de hibridación identificaron también los clones de ADNc con secuencias de intrón no procesadas, dado que mostraron una hibridación preferente a un conjunto de fragmentos de ADN genómico. Para los clones de ADNc confirmados, sus correspondientes segmentos de ADN genómico se aislaron y los exones y límites de exón/intrón se secuenciaron. Como se indica en la figura 7, se identificó un total de 24 exones. Basado en esta información y en los resultados de los experimentos de representación en mapa físicos, se estimó que el locus del gen abarcaba 250 kb en el cromosoma 7.

2.6 La secuencia

La figura 1 muestra la secuencia de nucleótidos del ADNc clonado que codifica CFTR junto con la secuencia de aminoácidos deducida. La primera posición de la base corresponde al primer nucleótido en el clon PA3-5 en la extensión 5' que es un nucleótido más largo que TB2-7. Las flechas indican la posición del sitio de iniciación de la transcripción mediante análisis de la extensión del cebador. Al nucleótido 6129 le sigue un tracto poli(dA). Mediante líneas verticales se indican las posiciones de las uniones del exón. Los segmentos potenciales que abarcan la membrana se verificaron utilizando el algoritmo de Eisenberg et al *J. Mol. Biol.* 179:125 (1984). En los recuadros de la figura 1 se incluyen segmentos potenciales que abarcan la membrana según se analizan y muestran en la figura 11. En la figura 11, se representa el índice medio de hidropatía [Kyte y Doolittle, *J. Molec. Biol.* 157, 165 (1982)] de 9 péptidos residuales frente al número de aminoácidos. Las posiciones correspondientes de rasgos de la estructura secundaria predicha de acuerdo con Garnier et al, [*J. Molec. Biol.* 157, 165 (1982)] se indican en el panel inferior. En la figura 1 están subrayados los aminoácidos que comprenden pliegues que unen ATP putativos. Posibles sitios de fosforilación mediante proteína quinasas A (PKA - protein kinases A) o C (PKC) se indican mediante círculos en blanco y en negro, respectivamente. El triángulo en blanco se encuentra por encima del (CTT) de 3 pb que se suprimen en CF (véase el comentario de más abajo). Los clones de ADNc en la figura 1 se secuenciaron mediante el método de terminación en cadena dideoxi empleando nucleótidos marcados con ³⁵S mediante el secuenciador automático de ADN Dupont Genesis 2000®.

La secuencia de ADNc combinada abarca 6129 pares de bases, excluyendo la cola poli(A) en el extremo de la región 3' no traducida y contiene un ORF capaz de codificar un polipéptido de 1480 aminoácidos (figura 1). Un triplete ATG (AUG) está presente al comienzo de este ORF (posición de la base 133-135). Dado que la secuencia de nucleótidos que rodea a este codón (5'-AGACCAUGCA-3') tiene las características propuestas de la secuencia consenso (CC) A/GCCAUGG(G) de un sitio de iniciación de la traducción eucariótica con un A muy conservado en la posición -3, es muy probable que este AUG corresponda al primer codón metionina para el polipéptido putativo.

Para obtener la secuencia correspondiente al extremo 5' transcrito, se realizó un experimento de extensión del cebador, según se describió anteriormente. Como se muestra en la figura 10A, se pudo observar un producto de extensión del cebador de aproximadamente 216 nucleótidos, sugiriendo que el extremo 5' del transcrito se iniciaba aproximadamente 60 nucleótidos más arriba del extremo del clon 10-1 de ADNc. A continuación se utilizó una reacción en cadena de la polimerasa modificada (PCR anclada) para facilitar la clonación de las secuencias en el extremo 5' (figura 10b). Dos clones de extensión 5' independientes, uno procedente del páncreas y el otro procedente de ARN de T84, se caracterizaron mediante secuenciación de ADN y se encontró que diferían en sólo 1 base en longitud, indicando el sitio de iniciación más probable para el transcrito según se muestra en la figura 1.

Dado que la mayoría de los clones de ADNc iniciales no contenían una cola poliA indicativa del extremo de un ARNm, la PCR anclada se aplicó también al extremo 3' del transcrito (Frohman et al, 1988, *supra*). Se realizaron tres oligonucleótidos en la extensión 3' a la porción terminal del clon T16-4.5 de ADNc. Como se muestra en la figura 10c, se obtuvieron 3 productos de PCR de tamaños diferentes. Todos ellos eran consistentes con la interpretación de que el extremo del transcrito se encontraba aproximadamente 1,2 kb más abajo del sitio Hind III en la posición 5027 del nucleótido (véase la figura 1). La secuencia de ADN derivada de los clones representativos estaba en concordancia con la del clon T12a de ADNc de T84 (véanse las figuras 1 y 7), y la secuencia del correspondiente fragmento genómico Eco RI de 2,3 kb.

3.0 Genética molecular de CF

3.1 Sitios de expresión

Para visualizar el transcrito para el gen de CF putativo, se llevaron a cabo experimentos de hibridación de borrones en gel de ARN con el ADNc de 10-1 como sonda. Los resultados de la hibridación de ARN se muestran en la figura 8.

Muestras de ARN se prepararon a partir de muestras de tejido obtenidas de patología quirúrgica o de autopsias, de acuerdo con los métodos previamente descritos (A.M. Kimmel, S.L. Berger, comps. *Meth. Enzymol.* 152, 1987). Geles de formaldehído se transfirieron sobre membranas de nilón (Zetaprobe®; BioRad Lab). A continuación, las membranas se hibridaron con sondas de ADN marcadas hasta una elevada actividad específica mediante el método de cebado aleatorio (A.P. Feinberg y B. Vogelstein, *Anal. Biochem.* 132, 6, 1983) de acuerdo con procedimientos previamente publicados (J. Rommens et al, *Am. J. Hum. Genet.* 43, 645-663, 1988). La figura 8 muestra la hibridación por parte del clon 10-1 de ADNc a un transcrito de 6,5 kb en los tejidos indicados. El ARN total (10 µg) de cada uno de los tejidos

ES 2 210 224 T3

y ARN de poli A+ (1 μ g) de la línea de células de carcinoma de colon T84 se separaron sobre un gel de formaldehído al 1%. Se indican las posiciones de las bandas de ARNr 28S y 18S. Las flechas indican la posición de los transcritos. La medición se estableció comparándolo con marcadores de ARN convencionales (BRL Labs). HL60 es una línea de células de leucemia promielocítica humana, y T84 es una línea de células de cáncer de colon humano.

El análisis revela una banda prominente de aproximadamente 6,5 kb de tamaño en células T84. De manera similar, también se detectaron fuertes señales de hibridación en el páncreas y en cultivos primarios de células procedentes de pólipos nasales, sugiriendo que el ARNm maduro del gen de CF putativo es de aproximadamente 6,5 kb. Señales de hibridación secundarias, que representan probablemente productos de degradación, se detectaron en los intervalos de tamaño inferiores, pero éstas variaban entre diferentes experimentos. Se obtuvieron resultados idénticos con otros clones de ADNc como sondas. Basado en la intensidad de la banda de hibridación y de la comparación con los detectados para otros transcritos en condiciones experimentales idénticas, se estimó que los transcritos de CF putativos constituían aproximadamente el 0,01% del ARNm total en células T84.

También se examinó un cierto número de otros tejidos mediante análisis de hibridación por borrones en gel de ARN en un intento de correlacionar el modelo de expresión del gen 10-1 y la patología de CF. Como se muestra en la figura 8, transcritos, todos de idéntico tamaño, se encontraron en los pulmones, colon, glándulas sudoríparas (células epiteliales cultivadas), placenta, hígado y glándula parótida, pero las intensidades de señal en estos tejidos variaban entre diferentes preparaciones y eran generalmente más débiles que la detectada en el páncreas y en los pólipos nasales. La intensidad variaba entre diferentes preparaciones, por ejemplo la hibridación en el riñón no se detectó en la preparación mostrada en la figura 8, pero se puede discernir en subsiguientes análisis repetidos. No se pudo discernir ninguna señal de hibridación en el cerebro ni glándula adrenal (figura 8), ni en líneas de células de fibroblastos y linfoblastos de la piel.

En resumen, la expresión del gen de CF pareció producirse en muchos de los tejidos examinados, con niveles mayores en aquellos tejidos gravemente afectados por CF. Mientras que este modelo de expresión específico del tejido epitelial está en buena concordancia con la patología de la enfermedad, no se ha detectado ninguna diferencia significativa en la cantidad ni tampoco en el tamaño de los transcritos procedentes de tejidos de CF y testigo, consistente con la presunción de que las mutaciones de CF son cambios sutiles al nivel de nucleótido.

3.2 La mutación principal de CF

La figura 17 muestra la secuencia de ADN en la delección F508. A la izquierda, el complemento reserva de la secuencia desde la posición de la base 1649-1664 de la secuencia normal (según se deriva del clon T16 de ADNc). La secuencia de nucleótidos se exhibe como el rendimiento (en unidades de intensidad de fluorescencia arbitraria, eje y) representada frente al tiempo (eje x) para cada uno de los 2 tubos fotomultiplicadores (PMT - photomultiplier tube n° 1 y n° 2) de un sistema de análisis de ADN Dupont Genesis 2000®. Debajo se muestra la correspondiente secuencia de nucleótidos. A la derecha se encuentra la misma región procedente de una secuencia mutante (según se deriva del clon C16 de ADNc). Se prepararon plantillas de ADN de plásmido de doble cadena mediante el procedimiento de lisis alcalina. Cinco μ g de ADN de plásmido y 75 ng de cebador de oligonucleótido se utilizaron en cada reacción de secuenciación de acuerdo con el protocolo recomendado por Dupont, excepto que la reasociación se realizó a 45°C durante 30 min, y que la etapa de elongación/terminación duró 10 min a 42°C. Los nucleótidos fluorescentes incorporados se separaron por precipitación del producto de reacción de secuenciación de ADN con etanol en presencia de acetato de amonio 2,5 M a pH 7,0 y se aclararon una vez con etanol al 70%. El cebador utilizado para la secuenciación de T16-1 era un oligonucleótido específico 5'GTTGGCATGCTTTGATGACGCTTC3' que se extiende desde la posición de la base 1708-1731 y que para C16-1 era el cebador universal SK para el vector Bluescript (Stratagene). La figura 18 muestra también la secuencia de ADN en torno a la delección F508, según se determina mediante secuenciación manual. La secuencia normal desde la posición de la base 1726-1651 (desde T16-1 de ADNc) se muestra junto a la secuencia de CF (desde C16-1 de ADNc). El panel de la izquierda muestra las secuencias desde las hebras codificadoras obtenidas con el cebador B (5'GTTTTCTGGAT-TATGCCTGGGCAC3'), y el panel de la derecha las de la hebra opuesta con el cebador D (5'GTTGGCATGCTTTGATGACGCTTC3'). Los corchetes indican los tres nucleótidos en la normal que están ausentes en CF (cabeceras de flecha). La secuenciación se realizó según se describe en F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Nat. Acad. Sci. USA* 74: 5463 (1977).

Para investigar la proporción de pacientes de CF que portan esta delección (F508), muestras de ADN genómico procedentes de pacientes y sus progenitores se amplificaron en cada caso con cebadores de oligonucleótidos que flanquean la mutación en una reacción en cadena de la polimerasa y se hibridaron oligonucleótidos marcados con ³²P específicos para las secuencias normal y mutante putativa (véase la figura 2). Los resultados de este análisis se muestran en la Tabla 2.

ES 2 210 224 T3

TABLA 2

Distribución de cromosomas de CF y no CF(N) con y sin la delección de 3 pb

5	a.		<u>Cromosomas de CF</u>	<u>Cromosomas N</u>
		sin la delección	69	198
10		<u>con la delección</u>	<u>145</u>	<u>0</u>
		Total	214	198
15	b.	<u>Cromosomas de CF</u>		
		<u>con la delección de 3 pb</u>	<u>sin la delección</u>	
		CF-PI	62	24
20		CF-PS	5	9
		no clasificados	<u>78</u>	<u>36</u>
		Total	145 (68%)	69 (32%)

25 Los datos para los cromosomas CF-PI (pancreáticos insuficientes) y CF-PS (pancreáticos suficientes) se derivaron de las familias de CF utilizadas en el análisis de enlace. Estas familias se seleccionaron originalmente sin el conocimiento en lo que se refiere a PI o PS; las 15 familias de CF-PS subsiguientemente identificadas no se incluyeron como parte de este cálculo. Los cromosomas de CF no clasificados se obtuvieron del laboratorio de diagnóstico del Hospital de Niños Enfermos en Toronto y para los cuales no estaban disponibles datos de la función pancreática.

30 Se puede observar que el 68% (145/214) de los cromosomas CF en la población de pacientes generales tenía la delección F508 (Tabla 2). En contraposición, ninguno (0/198) de los cromosomas N tenía la delección (Tabla 2; $\chi^2 = 207$, $p < 10^{-57.5}$), sugiriendo que esta alteración de la secuencia es específica para CF y que es la mutación principal la que provoca la enfermedad. No se ha detectado ninguna recombinación entre la delección F508 y CF.

35 Se observaron otras diferencias en la secuencia entre los clones de ADNc normales (T16-4.5) y CF (C1-1/5). En la posición de base 2629, T16-4.5 mostró una C y C1-1/5 tenía una T, dando como resultado un cambio de Leu a Phe al nivel de aminoácido. En la posición 4555, la base era G en T16-4.5, pero era A en C1-1/5 (de Val a Met). Se piensa que estos hallazgos representan el polimorfismo de la secuencia. Análisis de hibridación de oligonucleótidos específicos de ADN de paciente/familia identificarán a estos como otras posibles mutaciones. Se observaron diferencias adicionales en los nucleótidos en las regiones 3' no traducidas entre diferentes clones de ADNc y la secuencia de ADN genómico. Son posibles tales diferencias y, según se aprecia, otras modificaciones en la secuencia; por ejemplo, diferencias que son debidas a polimorfismos de la secuencia normal y artefactos de clonación, siendo todas estas diferencias esencialmente equivalentes a las secuencias según se describe en la figura 1 en términos de su función y de sus aplicaciones comerciales.

50 Los amplios datos de representación en mapa genético y físico han dirigido estudios de clonación molecular para concentrarse en un pequeño segmento de ADN en el cromosoma 7. Dada la carencia de delecciones y redistribuciones de cromosomas en CF y la carencia de un análisis funcional bien desarrollado para el producto génico de CF, la identificación del gen de CF requería una caracterización detallada del propio locus y una comparación entre los alelos de CF y normales (N). Individuos aleatorios, fenotípicamente normales, no pudieron incluirse como testigos en la comparación debido a la elevada frecuencia de portadores carentes de síntomas en la población. Como resultado, para el análisis sólo fueron adecuados progenitores de pacientes de CF, cada uno de los cuales, por definición, porta un cromosoma N y un cromosoma CF. Además, debido a la fuerte asociación alélica observada entre CF y algunos de los marcadores de ADN estrechamente relacionados, fue necesario excluir la posibilidad de que las diferencias en la secuencia detectadas entre N y CF eran polimorfismos asociados con el locus de la enfermedad.

3.3 Identificación de RFLPs y estudios de familia

60 Para determinar la relación de cada uno de los segmentos de ADN aislados a partir de los experimentos de marcha y salto del cromosoma con CF, polimorfismos de la longitud del fragmento de restricción (RFLPs - restriction fragment length polymorphisms) se identificaron y utilizaron para estudiar familias en las que previamente se habían detectado sucesos de cruzamiento entre marcadores de CF y otros marcadores de ADN flanqueantes. Como se muestra en la figura 14, en la región de 500 kb se detectó un total de 18 RFLPs; 17 de ellos (desde E6 hasta CE1.0) están listados en la Tabla 3; algunos de ellos corresponden a marcadores previamente reseñados.

Cinco de los RFLPs, a saber 10-1X.6, T6/20, H1.3 y CE1.0 se identificaron con sondas de ADNc y ADN genómico derivadas del gen de CF putativo. Los datos de RFLP se presentan en la Tabla 3, con marcadores en las regiones MET y D7S8 incluidos para fines comparativos. Las distancias físicas entre estos marcadores así como su relación con las regiones MET y D7S8 se muestran en la figura 14.

TABLA 3. RFLPs ASOCIADOS CON EL GEN DE CF

Nombre de sonda	Enzima	Long. fragmento	Nº de bandas	CE-PT-01	Δ (kb)	Δ (cM)	Referencia
metD	BanI	7,6 (kb)	28	48	0,60	0,10	J. E. Spence et al, Am. J. Hum. Genet. 39:729 (1986)
		6,8	59	25			
metD	TaqI	6,2	74	75	0,66	0,06	R. White et al, Nature 318:382 (1985)
		4,8	19	4			
metE	TaqI	7,5	45	49	0,35	0,05	White et al, Nature 318:382 (1985)
		4,0	38	20			
E6	TaqI	4,4	58	62	0,45	0,06	B. Keren et al, Am. J. Hum. Genet. 44:827 (1989)
		3,6	42	17			
E7	TaqI	3,9	40	16	0,47	0,07	
		3+0,9	51	57			

TABLA 3. (continuación)

PHL31	HinfI	0,4	81	33	0,73	0,15	J.M. Romera et al, <u>Ann. N. Y. Acad. Sci.</u> 43:645 (1988)
		0,3	18	47			
W301.4	HindIII	20	62	33	0,68	0,13	B. Kerem et al, <u>EMBO J.</u>
		10	22	47			
H2.2A	TaqI	2,1	39	53	0,64	0,09	X. Estivill et al, <u>Nature</u> 326:840 (1987); X. Estivill et al, <u>Genomics</u> 1:257 (1987)
(XV2C)		1,4	37	11			
EC1.4	HindIII	3,8	31	69	0,89	0,17	
		2,8	56	7			
EC1.4	EcoI	20	27	69	0,89	0,18	
		15	62	9			
JG261	PstI	7,8	69	10	0,88	0,18	X. Estivill et al <u>supra</u> y B. Kerem et al <u>supra</u>

TABLA 3. (continuación)

CE1.0	N2aI	5,5	81	73	0,41	0,03	
	4,7+0,8		8	3			
J32	SaCI	15	21	24	0,17	0,02	M.C. Larruzi et al et al , Genet. 44:695 (1969)
		6	47	38			
J3.11	MspI	4,2	36	38	0,29	0,04	B.J. Wainwright et al, Nature 318:384 (1985)
		1,8	62	36			
J29	PvuII	9	26	36	0,36	0,06	M.C. Larruzi et al, et al , EMBO
		6	55	36			

ES 2 210 224 T3

Notas para la tabla 3

(a) El número de cromosomas N y CF-PI (CF con insuficiencia pancreática) se derivó de los progenitores en las familias utilizadas en análisis de enlace [Tsui et al, Cold Spring Harbor Symp. Quant. Biol. 51:325 (1986)].

(b) Asociación estandarizada (A), que está menos influenciada por la fluctuación de la distribución del alelo marcador de ADN entre los cromosomas N, se utiliza aquí para la comparación del coeficiente de asociación de Yule $A=(ad-bc)/(ad+bc)$, en que a, b, c y d son el número de cromosomas N con el alelo marcador de ADN 1, CF con 1, N con 2 y CF con 2, respectivamente. El riesgo relativo se puede calcular utilizando la relación $RR = (1+A)/(1-A)$ o su inversa.

(c) Se incluye para fines comparativos la asociación alélica (*), calculada de acuerdo con A. Chakravarti et al, Am. J. Hum. Genet. 36:1239, (1984), asumiendo la frecuencia de 0,02 para cromosomas CF en la población.

Debido al pequeño número de familias recombinantes disponibles para el análisis, como se esperaba de la estrecha distancia entre los marcadores estudiados y CF y la posibilidad de un diagnóstico erróneo, fueron necesarios enfoques alternativos en una representación dina en mapa adicional del gen de CF.

3.4 Asociación alélica

La asociación alélica (desequilibrio de enlace) ha sido detectada para muchos marcadores de ADN estrechamente vinculados. Mientras que la utilidad de utilizar la asociación alélica para medir la distancia genética es incierta, se ha observado una correlación global entre CF y los marcadores de ADN flanqueantes. Se observó una fuerte asociación con CF para los marcadores de ADN más próximos, D7S23 y D7S122, mientras que se detectó una asociación pequeña o ninguna para los marcadores más distantes MET, D7S8 o D7S424 (véase la figura 1).

Como se muestra en la Tabla 3, el grado de asociación entre los marcadores de ADN y CF (medido por el coeficiente de asociación de Yule) aumentó desde 0,35 para meth y 0,17 para J32 hasta 0,91 para 10-1X.6 (en el análisis sólo se utilizaron familias de pacientes CF-PI, ya que parecían ser genéticamente más homogéneas que CF-PS). Los coeficientes de asociación parecieron ser más bien constantes a lo largo de las 300 kb desde EG1.4 hasta H1.3; la fluctuación detectada en varias localizaciones, lo más significativamente en H2.3A, E4.1 y T6/20, se debió probablemente a la variación en la distribución alélica entre los cromosomas N (véase la Tabla 2). Por lo tanto, estos datos son consistentes con el resultado del estudio de familias recombinantes (véase la figura 14). Una conclusión similar se pudo también sacar mediante la inspección de los haplotipos del marcador de ADN extendido asociados con los cromosomas CF (véase más abajo). Sin embargo, la fuerte asociación alélica detectada a lo largo de la gran distancia física entre EG1.4 y H1.3 no permitió una representación adicional refinada en mapa del gen de CF. Dado que J44 era el último clon de ADN genómico asilado mediante marcha y salto del cromosoma antes de que se identificara un clon de ADNc, la fuerte asociación alélica detectada para el intervalo JG2E1-J44 impulsó a los autores de la invención a investigar secuencias de genes candidatas por todo este intervalo. Es de interés señalar que el grado más elevado de asociación alélica se detectó, de hecho, entre CF y los 2 RFLPs detectados por 10-1X.6, una región próxima a la mutación principal de CF.

La Tabla 4 muestra una asociación alélica por pares entre marcadores de ADN estrechamente vinculados a CF. El número medio de cromosomas utilizado en estos cálculos era de 75-80 y únicamente se utilizaron cromosomas procedentes de familias CF-PI en la valoración de cromosomas de CF. Se obtuvieron resultados similares cuando se utilizó la asociación estandarizada de Yule (A).

(Tabla pasa a página siguiente)

CROMOSOMAS N

TABLA 3

	metD	metH	E6	E7	pH131	D1A	H2.3A	EG1.4	JG2E1	E2.6	H2.8	E4.1	J44	10-1X.6	T620	H1.3	CE1.0	J32	J3.11	J29			
metD Band	0.35	0.49	0.04	0.04	0.05	0.07	0.27	0.05	0.06	0.07	0.14	0.07	0.09	0.03	0.06	0.10	0.03	0.16	0.05	0.07	0.11	0.02	
metD TaqI	0.21	0.41	0.13	0.15	0.02	0.01	0.02	0.09	0.15	0.11	0.07	0.24	0.03	0.11	0.08	0.02	0.05	0.13	0.15	0.09	0.09	0.05	
metH TaqI	0.01	0.14	0.01	0.05	0.06	0.06	0.24	0.05	0.06	0.07	0.13	0.15	0.07	0.04	0.02	0.02	0.02	0.03	0.21	0.04	0.18		
E6 TaqI	0.11	0.39	0.00	0.00	0.03	0.07	0.04	0.02	0.03	0.00	0.19	0.02	0.09	0.19	0.09	0.11	0.09	0.15	0.07	0.11	0.20	0.00	
E7 TaqI	0.16	0.31	0.02	1.00	0.11	0.09	0.03	0.03	0.04	0.01	0.11	0.00	0.07	0.22	0.01	0.02	0.09	0.13	0.06	0.06	0.16	0.04	
pH131 Hind	0.45	0.28	0.23	0.38	0.48	0.01	0.12	0.04	0.09	0.05	0.05	0.03	0.03	0.08	0.16	0.15	0.20	0.04	0.03	0.06	0.08	0.06	
W3D1.4 Hind	0.45	0.28	0.23	0.45	0.47	0.35	0.21	0.02	0.03	0.01	0.05	0.03	0.03	0.10	0.12	0.10	0.23	0.10	0.05	0.05	0.10	0.06	
H2.3A TaqI	0.20	0.11	0.15	0.08	0.11	0.38	0.47	0.05	0.11	0.07	0.42	0.14	0.29	0.07	0.27	0.22	0.20	0.09	0.23	0.04	0.08	0.12	
EG1.4 Hind	0.11	0.08	0.07	0.05	0.07	0.20	0.20	0.24	0.05	0.07	0.75	0.05	0.01	0.00	0.07	0.13	0.01	0.56	0.04	0.24	0.14	0.15	
E4.1 Bgl	0.03	0.08	0.07	0.08	0.07	0.27	0.49	1.00	0.02	0.77	0.93	0.71	0.55	0.06	0.07	0.56	0.55	0.12	0.28	0.24	0.20	0.20	
JG2E1 PstI	0.07	0.08	0.03	0.09	0.08	0.30	0.30	0.34	0.01	0.02	0.92	0.04	1.00	0.76	0.04	0.11	0.11	0.61	0.57	0.13	0.31	0.26	0.22
E2.6/E.9 MspI	0.22	0.06	0.07	0.02	0.03	0.30	0.29	0.34	0.01	0.02	0.92	0.04	0.97	0.75	0.56	0.52	0.47	0.70	0.32	0.31	0.25	0.22	
H2.8 NcoI	0.05	0.07	0.01	0.08	0.06	0.31	0.31	0.45	0.02	0.93	1.00	0.92	0.74	0.63	0.13	0.18	0.69	0.59	0.10	0.28	0.28	0.18	
E4.1 MspI	0.12	0.06	0.07	0.05	0.03	0.25	0.48	0.02	0.05	0.94	1.00	0.93	0.71	0.49	0.49	0.49	0.68	0.35	0.27	0.25	0.21	0.18	
J44 XbaI	0.18	0.05	0.05	0.01	0.01	0.26	0.26	0.45	0.71	0.09	0.90	0.90	0.85	0.33	0.49	0.65	0.64	0.32	0.24	0.22	0.23	0.23	
10-1X.6 AccI	0.16	0.10	0.24	0.10	0.11	0.42	0.42	0.64	0.54	0.54	0.64	0.70	0.69	0.69	0.59	0.00	0.59	0.00	0.56	0.00	0.02	0.03	
10-1X.6 HaeIII	0.16	0.10	0.26	0.08	0.11	0.41	0.41	0.65	0.54	0.54	0.78	0.69	0.69	0.59	1.00	0.18	0.45	0.02	0.02	0.02	0.02	0.06	
T620 MspI	0.27	0.07	0.36	0.13	0.13	0.23	0.23	0.29	0.05	0.09	0.01	0.07	0.02	0.01	0.11	0.69	0.00	0.58	0.03	0.21	0.18	0.25	
H1.3 NcoI	0.06	0.06	0.05	0.03	0.01	0.30	0.30	0.56	0.71	0.78	0.07	0.90	0.92	0.94	0.94	0.12	0.40	0.19	0.13	0.20	0.14	0.20	
CE1.0 NcoI	0.00	0.04	0.02	0.11	0.11	0.25	0.25	0.08	0.09	0.55	0.43	0.53	0.37	0.44	0.24	0.24	0.07	0.40	0.19	0.20	0.14	0.14	
J32 SacI	0.03	0.13	0.07	0.17	0.13	0.17	0.24	0.07	0.21	0.24	0.22	0.24	0.21	0.21	0.27	0.26	0.13	0.21	0.18	0.04	0.97	0.97	
J3.11 MspI	0.14	0.11	0.15	0.07	0.06	0.05	0.05	0.12	0.11	0.10	0.13	0.18	0.19	0.15	0.20	0.28	0.24	0.14	0.07	0.91	0.71	0.71	
J29 PvuII	0.11	0.12	0.09	0.10	0.10	0.00	0.00	0.10	0.10	0.10	0.14	0.17	0.20	0.15	0.18	0.29	0.23	0.16	0.06	0.85	0.37	0.37	

CROMOSOMAS CF

5
10
15
20
25
30
35
40
45
50
55
60
65

ES 2 210 224 T3

También se detectó una fuerte asociación alélica entre subgrupos de RFLPs tanto sobre los cromosomas de CF como N. Como se muestra en la Tabla 4, los marcadores de ADN que están físicamente próximos uno con otro parecieron tener generalmente una asociación más intensa uno con otro. Por ejemplo, una fuerte asociación alélica (en algunos casos casi completa) se detectó entre marcadores adyacentes E6 y E7, entre pH131 y W3D1.4, entre los sitios polimórficos AccI y HaeIII detectados por 10-1X.6 y entre EG1.4, JG2E1, E2.6(E.9), E2.8 y E4.1. Los dos grupos de marcadores distales en la región MET y D7S8 también mostraron un cierto grado de desequilibrio de enlace entre ellos mismos, pero mostraron una pequeña asociación con marcadores desde E6 hasta CE1.0, consistente con las localizaciones distantes para MET y D7S8. Por otro lado, la carencia de asociación entre marcadores de ADN que son físicamente próximos puede indicar la presencia de puntos calientes de recombinación. Ejemplos de estos puntos calientes potenciales son la región entre E7 y pH131, en torno a H2.3A, entre J44 y las regiones cubiertas por las sondas 10-1X.6 y T6/20 (véase la figura 14). Estas regiones, que contienen frecuentes puntos de ruptura en la recombinación, eran útiles en el subsiguiente análisis de datos de haplotipo ampliados para la región de CF.

3.5 *Análisis del haplotipo*

Haplotipos ampliados basados en 23 marcadores de ADN se generaron para los cromosomas de CF y N en la colección de familias previamente utilizadas para el análisis de enlace. Asumiendo una recombinación entre cromosomas de diferentes haplotipos, fue posible construir varios linajes de los cromosomas de CF observados y, también, predecir la localización del locus de la enfermedad.

Para obtener una información adicional, útil para comprender la naturaleza de diferentes mutaciones de CF, los datos de la delección F508 se correlacionaron con los haplotipos del marcador de ADN ampliado. Como muestra la Tabla 5, cinco grupos principales de haplotipos N y de CF se pudieron definir por los RFLPs dentro de o inmediatamente junto al gen de CF putativo (regiones 6-8).

(Tabla pasa a página siguiente)

TABLA 5 HAPLOTIPOS DEL MARCADOR DE ADN QUE ABARCAN EL LOCUS DE CF

I. (a)	HAPLOTIPOS (a)		CF (b)			
	1	2	PI (F508)	PS (F508)	PI (F508)	PS (F508)
1	A	A	10	1	1	1
2	A	A	3	1	1	1
3	A	A	1	1	1	1
4	A	A	10	1	1	1
5	A	A	4	1	1	1
6	A	A	1	1	1	1
7	A	A	4	1	1	1
8	A	A	1	1	1	1
9	A	A	1	1	1	1
10	A	A	1	1	1	1
11	A	A	1	1	1	1
12	A	A	1	1	1	1
13	A	A	1	1	1	1
14	A	A	1	1	1	1
15	A	A	1	1	1	1
16	A	A	1	1	1	1
17	A	A	1	1	1	1
18	A	A	1	1	1	1
19	A	A	1	1	1	1
20	A	A	1	1	1	1
21	A	A	1	1	1	1
22	A	A	1	1	1	1
23	A	A	1	1	1	1
24	A	A	1	1	1	1
25	A	A	1	1	1	1
26	A	A	1	1	1	1
27	A	A	1	1	1	1
28	A	A	1	1	1	1
29	A	A	1	1	1	1
30	A	A	1	1	1	1
31	A	A	1	1	1	1
32	A	A	1	1	1	1
33	A	A	1	1	1	1
34	A	A	1	1	1	1
35	A	A	1	1	1	1
36	A	A	1	1	1	1
37	A	A	1	1	1	1
38	A	A	1	1	1	1
39	A	A	1	1	1	1
40	A	A	1	1	1	1
41	A	A	1	1	1	1
42	A	A	1	1	1	1
43	A	A	1	1	1	1
44	A	A	1	1	1	1
45	A	A	1	1	1	1
46	A	A	1	1	1	1
47	A	A	1	1	1	1
48	A	A	1	1	1	1
49	A	A	1	1	1	1
50	A	A	1	1	1	1
51	A	A	1	1	1	1
52	A	A	1	1	1	1
53	A	A	1	1	1	1
54	A	A	1	1	1	1
55	A	A	1	1	1	1
56	A	A	1	1	1	1
57	A	A	1	1	1	1
58	A	A	1	1	1	1
59	A	A	1	1	1	1
60	A	A	1	1	1	1
61	A	A	1	1	1	1
62	A	A	1	1	1	1
63	A	A	1	1	1	1
64	A	A	1	1	1	1
65	A	A	1	1	1	1

5
10
15
20
25
30
35
40
45
50
55
60
65

	1 1 1 1 1	14
 1 1 1 1	1
 1 1 1 1	7
 1 1 1 1	5
 1 1 1 1	57
	A B A A B B B B B B C A A B	
	. A/D A A A A A A D D D D D D D D C C	
	A A A A A A A A A A A A . . A A	
	A A A A A A A A A A A A A A A A	
	A . A A A A A C C C C C C C A A A	
	B A A A A A B B B B B B B B A C	
	A B . A . A . A A A A A A A A B	
	B C C C C C C C C C A B B D B B C	
	A B A A D D F B B B Y Y A B B A A D B	
(b)		
(c)		
(d)		

 1 1 1 1	
 1 1 1 1	
 1 1 1 1	
 1 1 1 1	
 1 1 1 1	
	B B B B A A A A B C C B A D	
	C C A/C C C C C C C C C C C C C	
	A A A A A . A A . A A A A	
	B B B B B B B B B B B B B B	
	B B . B B B B B B B . B B B	
	B B B B B B B B B B B B B B	
	. B . B B B B B B . B B B B	
	A B/C A B B C C C C C C C C C C	
	B . B A B A A Y A A B B B B	
II.(a)		

ES 2 210 224 T3

Tabla 5 (continuación)

(a) Los datos de los haplotipos extendidos se derivan de las familias de CF utilizadas en estudios de linaje previos (véase la nota al pie de página (a) de la Tabla 3) con familias de CF-PS adicionales recogidas subsiguientemente (Kerem et al, *Am. J. Genet.* 44:827 (1989)). Los datos se muestran en grupos (regiones) para reducir espacio. Las regiones se asignan principalmente de acuerdo con los datos de asociación por parejas mostrados en la Tabla 4, abarcando las regiones 6-8 el locus de CF putativo (la delección F508 se encuentra entre las regiones 6 y 7). Se muestra un guión (-) en la región en la que no se ha determinado el haplotipo debido a datos incompletos o a imposibilidad de establecer una fase. También se dan asociaciones de haplotipo alternativas en los casos en los que los datos sean incompletos. "No clasificado" incluye los cromosomas con más de 3 asociaciones desconocidas. Las definiciones del haplotipo para cada una de las 9 regiones son:

15	Región 1-	metD	metD	methH	
	<u>BanI</u>	<u>TaqI</u>	<u>TaqI</u>		
	A =	1	1	1	
	B =	2	1	2	
20	C =	1	1	2	
	D =	2	2	1	
	E =	1	2	-	
	F =	2	1	1	
25	G =	2	2	2	
30	Región 2-	E6	E7	pH131	W3D1.4
	<u>TaqI</u>	<u>TaqI</u>	<u>HinfI</u> <u>HindIII</u>		
	A =	1	2	2	2
	B =	2	1	1	1
35	C =	1	2	1	1
	D =	2	1	2	2
	E =	2	2	2	1
	F =	2	2	1	1
40	G =	1	2	1	2
	H =	1	1	2	2
45	Región 3-	H2.3A			
	<u>TaqI</u>				
	A =	1			
50	B =	2			
55	Región 4-	EG1.4	EG1.4	JG2E1	
	<u>HincII</u>	<u>BglI</u>	<u>PstI</u>		
	A =	1	1	2	
	B =	2	2	1	
	C =	2	2	2	
60	D =	1	1	1	
	E =	1	2	1	
65					

ES 2 210 224 T3

TABLA 5 (continuación)

5	<u>Región 5-</u>	<u>E2.6</u>	<u>E2.8</u>	<u>E4.1</u>
	<u>MspI</u>	<u>NcoI</u>	<u>MspI</u>	
	A =	2	1	2
10	B =	1	2	1
	C =	2	2	2
15	<u>Región 6-</u>	<u>J44</u>	<u>10-1X.6</u>	<u>10-1X.6</u>
	<u>XbaI</u>	<u>AccI</u>	<u>HaeIII</u>	
	A =	1	2	1
	B =	2	1	2
20	C =	1	1	2
	D =	1	2	2
	E =	2	2	2
	F =	2	2	1
25	<u>Región 7-</u>	<u>T6/20</u>		
	<u>MspI</u>			
30	A =	1		
	B =	2		
35	<u>Región 8-</u>	<u>H1.3</u>	<u>CE1.0</u>	
	<u>NcoI</u>	<u>NdeI</u>		
	A =	2	1	
40	B =	1	2	
	C =	1	1	
	D =	2	2	
45	<u>Región 9-</u>	<u>J32</u>	<u>J3.11</u>	<u>J29</u>
	<u>SacI</u>	<u>MspI</u>	<u>PvuII</u>	
	A =	1	1	1
50	B =	2	2	2
	C =	2	1	2
	D =	2	2	1
55	E =	2	1	1

(b) Número de cromosomas anotado en cada clase:

60	CF-PI(F)	=	cromosomas de CF procedentes de pacientes CF-PI con la delección F508;
	CF-PS(F)	=	cromosomas de CF procedentes de pacientes CF-PS con la delección F508;
	CF-PI	=	otros cromosomas de CF procedentes de pacientes CF-PI;
	CF-PS	=	otros cromosomas de CF procedentes de pacientes CF-PS;
65	N	=	cromosomas normales derivados de progenitores portadores

Resultó evidente que la mayoría de las recombinaciones entre haplotipos se producían entre las regiones 1 y 2 y entre las regiones 8 y 9, de nuevo en buena concordancia con la distancia física relativamente larga entre estas regiones. Otros puntos de ruptura, menos frecuentes, se observaron entre intervalos de distancia corta, y éstos correspondían generalmente a los puntos calientes identificados por los estudios de asociación alélica por parejas como se muestra anteriormente. El resultado llamativo fue que la delección F508 se asociaba casi exclusivamente con el Grupo I, el haplotipo de CF más frecuente, lo que sustenta la postura que esta delección constituye la mutación principal en CF. De modo más importante, mientras que la delección F508 se detectó en el 89% (62/70) de los cromosomas de CF con el haplotipo AA (correspondiente a las dos regiones, 6 y 7), que flanquean la delección, no se encontró ninguna en los 14 cromosomas N dentro del mismo grupo ($\chi^2 = 47,3$, $p < 10^{-4}$). Por lo tanto, la delección F508 no era un polimorfismo de la secuencia común asociado al núcleo del haplotipo del Grupo I (véase la Tabla 5).

Se encontró que uno de los cromosomas de CF, detectado por la sonda de oligonucleótidos específica para la delección F508, pertenecía a un grupo de haplotipo diferente (Grupo III). Ninguno de los otros 9 cromosomas de CF ni de los 17 cromosomas N con el mismo grupo se hibridaba a la sonda. Este resultado de hibridación específico sugiere que la mutación albergada en este cromosoma es similar a F508. Aunque la recombinación o la conversión del gen son mecanismos posibles para explicar la procedencia de esta delección en un haplotipo de no Grupo I, es más probable que estos 2 cromosomas del Grupo III representen un suceso de mutación recurrente, una situación similar a las mutaciones β^S y β^E en el locus de β -globina.

Juntos, los resultados del estudio de hibridación de oligonucleótidos y el análisis del haplotipo sustentan el hecho de que el locus del gen descrito en esta memoria es el gen de CF y que la delección de 3 pb (F508) es la mutación más común en CF.

3.6 Otras mutaciones de CF

La asociación de la delección F508 con 1 haplotipo de CF común y 1 raro proporcionó una perspectiva adicional en el número de sucesos de mutación que podrían contribuir a la presente población de pacientes. En base a los amplios datos de haplotipo, es probable que los 2 cromosomas originales en los que se producía la delección F508 portaran el haplotipo -AAAAAAA- (Grupo Ia) y -CBAACBA- (Grupo IIIa), según se define en la Tabla 5. Los otros cromosomas de CF del Grupo I que portan la delección son probablemente productos de recombinación derivados del cromosoma original. Si se considera que los cromosomas de CF en cada grupo de haplotipo se derivan del mismo origen, únicamente se podrían predecir 3-4 sucesos de mutación adicionales (véase la Tabla 5). Sin embargo, dado que muchos de los cromosomas de CF en el mismo grupo son acusadamente diferentes uno de otro, es posible una subdivisión adicional dentro de cada grupo. Como resultado, se podría considerar un mayor número de sucesos independientes de mutación, y los datos sugieren que al menos 7 mutaciones adicionales putativas contribuyen también al fenotipo CF-PI (véase la Tabla 4). Las mutaciones que conducen al subgrupo CF-PS son probablemente más heterogéneas.

Las 7 mutaciones de CF-PI adicionales se representan por los haplotipos: -CAAAAAA- (Grupo Ib), -CABCAAD- (Grupo Ic), - - -BBBAC (Grupo IIa), -CABBBAB- (Grupo Va). A pesar de que el defecto molecular en cada una de estas mutaciones está todavía por definir, es claro que ninguna de estas mutaciones afecta gravemente a la región correspondiente a los sitios de unión del oligonucleótido utilizados en el experimento de hibridación por PCR.

3.7 Suficiencia pancreática

CF-PS se define clínicamente como función exocrina pancreática suficiente para la digestión de alimentos; sin embargo, el nivel de la actividad residual de enzimas pancreáticas en el sistema digestivo varía de un paciente a otro. Datos de haplotipo previos sugirieron que los pacientes de CF-PI y CF-PS son debidos a diferentes alelos mutantes. A pesar de que el defecto bioquímico básico en CF está todavía por definir, es posible que la actividad residual de enzima pancreática en pacientes CF-PS sea un reflejo directo de la actividad del producto génico de CF mutante. Así, la función exocrina residual conferida por un alelo suave (CF-PS), a pesar de que es mucho menor que la del producto génico normal, constituía un fenotipo dominante frente al de mutaciones más severas (CF-PI) con una pequeña función o sin ninguna. Se deduce que únicamente pacientes que portan 2 copias de alelos graves serían CF-PI y que los pacientes que portan 1 ó 2 alelos suaves serían CF-PS.

Para someter a ensayo la hipótesis anterior, se podría utilizar la información sobre la proporción de pacientes de CF portadores de la delección F508. Asumiendo que una mutación severa sea recesiva a una mutación suave y que una distribución de alelos de CF en la población de pacientes de acuerdo con la ley de Hardy-Weinberg, se podría estimar la frecuencia de alelos severos en 0,92 y la de alelos suaves (M), en 0,08 (véase la Tabla 6).

ES 2 210 224 T3

TABLA 6

Análisis de población de CF-PI y CF-PS

<u>do</u> ^(d)	Genotipo	Frecuencia	<u>Observado</u> ^(c)	<u>Espera-</u>	
	<u>asumido</u> ^(a)	<u>predicha</u> ^(b)			
10	Pancreático	FF	0,459	21	21,1
	insuficiente (PI)	FS	0,331	14	15,2
15		SS	0,060	4	2,7
	Total		0,850	39	-
20	Pancreático	FM	0,106	15 ^(e)	14,8
	suficiente (PS)	SM	0,038	6	6,2
		MM	0,006		
25		Total	0,150	21	

(a) Designaciones de los alelos: F = delección de 3 pb (delección de fenilalanina en la posición del aminoácido 508); S = alelos mutantes severos no caracterizados; M = alelos mutantes suaves no caracterizados.

(b) Asumiendo que el fenotipo mutante CF-PI sea recesivo al fenotipo mutante CF-PS, la frecuencia de alelos mutantes CF-PI, incluida la delección de 3 pb, podría estimarse a partir de la proporción observada de los pacientes CF-PI en la clínica de CF [Corey et al. *J. Pediatr.* 115:274 (1989)], es decir $(0,85) = 0,92$. La frecuencia de alelos observada para F en la población de CF total es 0,68 (Tabla 3); la frecuencia para S = $0,92 - 0,68 = 0,24$; la frecuencia para M = $1 - 0,92 = 0,08$. Luego se calculó la frecuencia para cada genotipo utilizando la ley de Hardy-Weinberg.

(c) El número de pacientes CF-PI y CF-PS en cada categoría se obtuvo mediante análisis de hibridación de oligonucleótidos como se ilustra en la figura 15. Los pacientes procedían de las familias de CF utilizadas en el análisis de enlace llevado a cabo por los autores de la invención con 14 pacientes/familias de CF-PS adicionales a partir de un estudio subsiguiente. Dado que SM y MM no podrían distinguirse genotípicamente ni fenotípicamente, éstos se combinaron en el análisis.

(d) Los números esperados se calcularon para CF-PI y CF-PS después de la normalización con cada grupo. La χ^2 de ajuste es 0,86, d.f. = 3, $0,74 < p < 0,90$.

(e) Este número es mayor que el que cabría esperar (15 observados frente a 9,6 esperados), si la delección F508 se encuentra en el equilibrio de Hardy-Weinberg en todos los cromosomas CF ($\chi^2 = 6,48$, d.f. = 1, $p < 0,011$).

Dado que se encontró que la mayoría de pacientes CF-PI eran homocigóticos para la mutación F508 (F), era razonable asumir que esta mutación correspondía a uno de varios alelos severos. Dada la frecuencia observada de F (0,68) en la población de CF estudiada, se podría derivar la frecuencia de los alelos severos (S) remanentes. Luego se calculó la proporción de pacientes FF, SS, MM, FS, FM y SM. Dado que individuos con SM y MM no podrían distinguirse fenotípicamente ni genotípicamente, éstos se combinaron en el análisis. Como se muestra en la Tabla 6, las frecuencias observadas para los 5 grupos de pacientes eran como las esperadas a partir de esta hipótesis.

El análisis anterior proporciona así un fuerte soporte de la posición de los autores de la invención de que CF-PI se debe a la presencia de 2 alelos severos y que un paciente de CF-PS porta un único alelo severo o 2 alelos suaves. Este modelo explica también la menor frecuencia de la delección F508 en la población CF-PS que en la CF-PI y el número en exceso de pacientes de CF-PS con una copia de la delección (véase la nota en la Tabla 6).

Dado el fenotipo dominante predicho conferido por los alelos M, fue necesario examinar los cromosomas de CF en pacientes CF-PS individualmente, con el fin de identificar los que portaran los elementos M. Tal como se muestra en la Tabla 7, cinco de los 7 pacientes CF-PS representativos portan una copia de la delección F508; se pudieron asignar al menos 5 haplotipos diferentes a los otros cromosomas de CF.

ES 2 210 224 T3

TABLA 7

Haplotipos de cromosomas de CF en individuos CF-PS y familias con MI

5

Familia nº 1 2 3 4 5 6 7 8 9 Alelos de CF

10

(a) individuos CF-PS

15

20

25

30

35

40

45

50

3	A	A	A	A	A	A	-	-	A	F (Grupo Ia)
	D	C	B	A	A	C	B	A	A	M (predicho, Grupo IIIb)
14	B	A	A	A	-	A	A	A	A	F (Grupo Ia)
	B	C	B	B	-	B	A	A	C	M (predicho, Grupo IIa)
27	A	B	-	A	A	A	A	A	E	F (Grupo Ia)
	A	C	A	A	A	A	A	A	A	M (predicho, Grupo Ib)
29	A	C	-	C	-	A	A	A	B	F (Grupo Ia)
	B	A	-	B	-	B	A	A/C	B	M (predicho, Grupo IIa)
40	D	A	A	A	A	A	A	A	B	F (Grupo Ia)
	F	C	B	A	A	C	B	C	A	M (predicho, Grupo IV)
51	C	C	A	B	B	B/C	A	C	C	M (predicho, Grupo IIa)
	F	D	A	B	B	B/C	A	C	C	M (predicho, Grupo IIa)
54	B	C	A	B	C	C	A	C	A	M o S (predicho, Grupo Vb)
	B	B	A	A	A	C	B	A	A	M (predicho, Grupo IIIb)

55

(b) familias con MI

60

65

4	B	A	A	A	A	A	A	A	A	F (Grupo Ia)
	B	A	A	A	A	A	A	A	A	F (Grupo Ia)
10	D	B	A	A	-	A	-	A	A	F (Grupo Ia)
	A	D	A	A	-	A	A	A	B	F (Grupo Ia)
23	A	E	B	A	A	A	A	A	E	F (Grupo Ia)
	B	C	A	A	A	A	A	A	B	S (predicho, Grupo Ib)

ES 2 210 224 T3

28	A	A	-	A	A	A	A	A	C	F (Grupo Ia)
	A	A	-	A	A	A	A	A	B	F (Grupo Ia)
33	B	B	-	A	A	A	-	A	B	F (Grupo Ia)
	B	A	-	A	A	A	A	A	B	F (Grupo Ia)
49	A	A	A	A	A	A	A	A	B	F (Grupo Ia)
	A	A	A	A	A	A	A	A	B	F (Grupo Ia)

(a) Las definiciones del haplotipo son las mismas que en la Tabla 5.

(b) Las designaciones de los alelos son las mismas que en la Tabla 6: F = la delección F508; S = alelo mutante severo no caracterizado; M = alelo mutante suave no caracterizado.

Estas últimas observaciones proporcionan un soporte adicional de que la mayoría de los pacientes CF-PS son heterocigotos compuestos.

4.0 Proteína CFTR

Como se ha comentado con respecto a la secuencia de ADN de la figura 1, el análisis de la secuencia de los clones de ADNc solapantes predijo un polipéptido no procesado de 1480 aminoácidos con una masa molecular de 168.138 dalton. Como se describe más tarde, debido a los polimorfismos en la proteína, el peso molecular de la proteína puede variar debido a las posibles sustituciones o delecciones de determinados aminoácidos. El peso molecular también cambiará debido a la adición de unidades de hidrato de carbono para formar un glicoproteína. Se entiende que la proteína funcional en la célula será similar al polipéptido no procesado, pero puede modificarse debido al metabolismo de la célula.

Por consiguiente, la invención proporciona polipéptido CFTR normal purificado caracterizado por un peso molecular de aproximadamente 170.000 dalton y con una actividad de la conductancia de iones de la membrana de las células epiteliales. El polipéptido CFTR normal, que está esencialmente exento de otras proteínas humanas, es codificado por las secuencias de ADN antes mencionadas y, de acuerdo con una realización, la de la figura 1. Un polipéptido de este tipo exhibe la actividad inmunológica o biológica de polipéptido CFTR normal. Como se discutirá más tarde, el polipéptido CFTR y fragmentos del mismo se pueden preparar mediante síntesis química o enzimática de péptidos o se puede expresar en un sistema de células cultivadas apropiado. La invención proporciona también un polipéptido CFTR mutante purificado que se caracteriza por una actividad asociada a la fibrosis quística en células epiteliales humanas. Un polipéptido CFTR mutante de este tipo, al estar sustancialmente libre de otras proteínas humanas, puede ser codificado por la secuencia de ADN mutante.

4.1 Estructura de CFTR

El rasgo más característico de la proteína predicha es la presencia de dos motivos repetidos, cada uno de los cuales consiste en un conjunto de residuos aminoácidos capaces de abarcar la membrana varias veces, seguido de una secuencia que se asemeja a pliegues de unión de nucleótidos (ATP) consenso (NBFs - nucleotide binding folds) (figuras 11, 12 y 16). Estas características son acusadamente similares a las de la glicoproteína P resistente a múltiples fármacos de mamíferos y a un cierto número de otras proteínas asociadas a la membrana, implicando de este modo que el producto del gen de CF predicho esté probablemente implicado en el transporte de sustancias (iones) a través de la membrana y sea probablemente un miembro de una superfamilia de proteínas de la membrana.

La figura 13 es un modelo esquemático de la proteína CFTR predicha. En la figura 13, los cilindros indican hélices que abarcan las membranas y las esferas rayadas indican NBFs. La esfera punteada es el dominio R polar. Las 6 hélices que abarcan las membranas en cada mitad de la molécula se describen como cilindros. Los NBFs citoplásmicamente orientados interiores se muestran en forma de esferas rayadas con rendijas para indicar los medios de entrada por parte del nucleótido. El gran dominio R polar que une las dos mitades representa una esfera punteada. Aminoácidos individuales cargados dentro de los segmentos de membrana y sobre la superficie del dominio R se describen como pequeños círculos que contienen el signo de carga. Las cargas netas en los bucles internos y externos que unen los cilindros de la membrana y regiones de los NBFs están contenidas en los cuadrados en blanco. Los sitios para la fosforilación por parte de proteína quinasas A o C se muestran mediante triángulos en negro y en blanco, respectivamente. K, R, H, D y E son una nomenclatura convencional para los aminoácidos lisina, arginina, histidina, ácido aspártico y ácido glutámico, respectivamente.

Cada una de las regiones asociadas a la membrana predichas de la proteína CFTR consiste en 6 segmentos muy

hidrófobos capaces de abarcar una bicapa de lípidos de acuerdo con los algoritmos de Kyte y Doolittle y de Garnier et al (J. Mol. Biol. 120, 97 (1978) (figura 13). Las regiones asociadas a la membrana están seguidas cada una por una gran región hidrófila que contiene los NBFs. En base a la alineación de la secuencia con otras proteínas de unión a nucleótidos conocidas, cada uno de los NBFs putativos en CFTR comprende al menos 150 residuos (figura 13). La

5 deleción de 3 pb detectada en la mayoría de pacientes CF está situada entre los 2 segmentos sumamente conservados del primer NBF en CFTR. La identidad en la secuencia de aminoácidos entre la región que rodea la deleción de fenilalanina y las correspondientes regiones de un cierto número de otras proteínas sugiere que esta región es de una importancia funcional (figura 16). Un aminoácido hidrófobo, habitualmente uno con una cadena lateral aromática, está presente en la mayoría de estas proteínas en la posición correspondiente a F508 de la proteína CFTR. Se entiende que

10 los polimorfismos de aminoácidos pueden existir como resultado de polimorfismos de ADN.

La figura 16 muestra la alineación de los 3 segmentos más conservados de los NBFs extendidos de CFTR con regiones equiparables de otras proteínas. Estos 3 segmentos consisten en los residuos 433-473, 488-513 y 542-584 de la mitad N-terminal, y 1219-1259, 1277-1302 y 1340-1382 de la mitad C-terminal de CFTR. El trazo grueso señala las

15 regiones de mayor similitud. Una homología general adicional puede observarse incluso sin la introducción de huecos.

A pesar de la simetría global en la estructura de la proteína y la conservación de la secuencia de los NBFs, la homología de la secuencia entre las dos mitades de la proteína CFTR predicha es modesta. Esto se demuestra en la figura 12, en la que los aminoácidos 1-1480 están representados sobre cada eje. Las líneas a ambos lados de la

20 diagonal de identidad indican las posiciones de similitudes internas. Por lo tanto, mientras que pueden detectarse cuatro conjuntos de identidad de secuencia interna como se muestra en la figura 12, al utilizar la matriz de registro de Dayhoff, según se aplica por Lawrence et al. [C. B. Lawrence, D. A. Goldman y R. T. Hood, Bull. Math. Biol. 48, 569 (1986)], únicamente tres de estos resultan evidentes a bajos ajustes del umbral para la desviación estándar. La identidad más intensa se encuentra entre secuencias en los extremos carboxilo de los NBFs. De los 66 residuos alineados, el 27%

25 es idéntico y otro 11% es funcionalmente similar. La homología interna débil global está en contraposición con el grado mucho mayor (> 70%) en la glicoproteína P para la cual se ha propuesto una hipótesis de duplicación del gen (Gros et al, Cell 47, 371, 1986, C. Chen et al, Cell 47, 381, 1986, Gerlach et al, Nature, 324, 485, 1986, Gros et al, Mol. Cell. Biol. 8, 2770, 1988). La carencia de conservación en las posiciones relativas de los límites exón-intrón pueden argumentar en contra de un modelo de este tipo para CFTR (figura 2).

30

Dado que aparentemente no existe ninguna secuencia de péptido señal en el extremo amino de CFTR, el segmento hidrófilo muy cargado que precede a la primera secuencia de la transmembrana está probablemente orientado en el citoplasma. Se espera que cada uno de los 2 conjuntos de hélices hidrófobas forme 3 bucles transversales a través de la

35 membrana, y se espera que una pequeña cantidad de la secuencia de la proteína completa esté expuesta a la superficie exterior, excepto la región entre los segmentos 7 y 8 de la transmembrana. Es de interés señalar que esta última región contiene dos sitios potenciales para la glicosilación N-enlazada.

A cada una de las regiones asociadas a la membrana le sigue un NBF según se indica anteriormente. Además, el dominio citoplásmico muy cargado puede identificarse en el centro del polipéptido CFTR predicho, que enlaza las dos

40 mitades de la proteína. Este dominio, denominado dominio R, está operativamente definido por un gran exón único en el que 69 de los 241 aminoácidos son residuos polares dispuestos en racimos alternantes en cargas positivas y negativas. Además, 9 de las 10 secuencias consenso requeridas para la fosforilación por parte de proteína quinasa A (PKA) y 7 de los potenciales sitios del sustrato para la proteína quinasa C (PKC) encontradas en CFTR están situadas

45 en este exón.

4.2 Función de CFTR

Las propiedades de CFTR se pueden derivar de una comparación con otras proteínas asociadas a la membrana (figura 16). Además de la similitud estructural global con la glicoproteína P de mamíferos, cada uno de los dos dominios

50 predichos en CFTR muestra también una semejanza notoria con la estructura del dominio único de hemolisina B de *E. coli* y el producto del gen Blanco de *Drosophila*. Estas últimas proteínas están implicadas en el transporte del péptido lítico del sistema hemolisina y de las moléculas de los pigmentos oculares, respectivamente. El sistema de transporte de la vitamina B12 de *E. coli*, BtuD y MbpX, que es un gen del cloroplasto hepático cuya función es desconocida, tienen también un motivo estructural similar. Además, la proteína CFTR comparte una similitud estructural

55 con varios de los sistemas de transporte de soluto periplásmicos de bacterias Gram-negativas en la región de la transmembrana y los pliegues que se unen a ATP están contenidos en proteínas separadas que funcionan acordes con un tercer polipéptido que se une al sustrato.

La disposición estructural global de los dominios de la transmembrana en CFTR es similar a varias proteínas del canal de cationes y a algunas ATPasas de translocaciones de cationes, así como a la adenilato ciclasa del cerebro bovino descrita recientemente. La importancia funcional de esta clasificación topológica, consistente en 6 dominios de la transmembrana, permanece siendo especulativa.

60

También se han detectado regiones cortas de identidad de secuencia entre las regiones de la transmembrana putativas de CFTR y otras proteínas que abarcan la membrana. De modo interesante, existen también secuencias, de 18 aminoácidos de longitud situadas aproximadamente 50 residuos del extremo carboxilo de CFTR y el proto-oncogen

65 raf de serina/treonina quinasa de *Xenopus laevis*, que son idénticos en 12 de estas posiciones.

Finalmente, se ha señalado una identidad de secuencia de aminoácidos (10/13 residuos conservados) entre un segmento hidrófilo (posición 701-713) dentro del dominio R muy cargado de CFTR y una región que precede inmediatamente al primer bucle de la membrana de los canales de sodio, tanto en el cerebro de ratas como en la anguila. El dominio R característico de CFTR no se comparte con la glicoproteína P estrechamente relacionada desde un punto de vista topológico; el péptido que enlaza 241 aminoácidos es aparentemente la diferencia principal entre las dos proteínas.

En resumen, los rasgos de la estructura primaria de la proteína CFTR indican su posesión de propiedades adecuadas para la participación en la regulación del control del transporte de iones en las células epiteliales afectados por CF. La fijación segura a la membrana en dos regiones sirve para posicionar sus tres dominios intracelulares principales (pliegues 1 y 2 de unión a nucleótidos y el dominio R) próximos a la superficie del citoplasma de la membrana celular en la que pueden modular el movimiento de iones a través de canales formados mediante segmentos de la membrana de CFTR por sí mismos o por otras proteínas de la membrana.

A la vista de los datos genéticos, la especificidad del tejido y las propiedades predichas de la proteína CFTR, es razonable concluir que CFTR es directamente responsable de la CF. Sin embargo, permanece siendo incierto el modo en que CFTR está implicada en la regulación de la conductancia de iones a través de la membrana apical de células epiteliales.

Es posible que CFTR sirva como un canal de iones por sí misma. Según se describe en la figura 13, 10 de las 12 regiones de la membrana contienen uno o más aminoácidos con cadenas laterales cargadas, una propiedad similar al canal del sodio del cerebro y las subunidades del canal de cloruro del receptor GABA, en que residuos cargados están presentes en 4 de los 6 y 3 de los 4 respectivos dominios asociados a la membrana por subunidad o unidad repetitiva. Se piensa que la naturaleza anfipática de estos segmentos de la membrana contribuye a la capacidad formadora de canales de estas moléculas. Alternativamente, CFTR puede no ser un canal de iones, sino, en su lugar, puede servir para regular las actividades del canal de iones. En apoyo de esta última presunción, ninguno de los polipéptidos purificados procedentes de la tráquea y los riñones son capaces de reconstituir canales de cloruro en las membranas de lípidos [Landry et al, *Science* 224:1469 (1989)] parecen ser CFTR si se juzgan sobre la base de la masa molecular.

En cualquier caso, la presencia de dominios que unen ATP en CFTR sugiere que la hidrólisis de ATP esté directamente implicada y es requerida para la función de transporte. La alta densidad de sitios de fosforilación para PKA y PKC y los racimos de residuos cargados en el dominio R pueden servir ambos para regular esta actividad. La delección de un residuo fenilalanina en el NBF puede prevenir una unión adecuada ATP o el cambio conformacional que esto produce normalmente y, por consiguiente, da como resultado la insensibilidad observada a la activación por parte de la fosforilación mediada por PKA o PKC de la vía de conductancia de cloruro apical de CF. Dado que la proteína predicha contiene varios dominios y pertenece a una familia de proteínas que frecuentemente funcionan como partes de sistemas moleculares multicomponentes, CFTR puede también participar en las funciones del tejido epitelial de actividad o regulación no relacionadas con el transporte de iones.

Con el gen de CF aislado (ADNc), ahora en posesión, es posible definir el defecto bioquímico básico en CF y elucidar adicionalmente el control de las vías de transporte de iones en las células epiteliales en general. Lo más importante, el conocimiento adquirido hasta ahora a partir de la estructura predicha de CFTR, junto con la información adicional de estudios de la propia proteína proporciona una base para el desarrollo de medios mejorados de tratamiento de la enfermedad. En este tipo de estudios, han surgido anticuerpos contra la proteína CFTR según se describe posteriormente.

4.3 Purificación de la proteína

La proteína CFTR se puede purificar por métodos seleccionados sobre la base de propiedades según se revelan por su secuencia. Por ejemplo, ya que posee distintas propiedades de una proteína de la membrana integral, se aísla primeramente una fracción de la membrana de las células epiteliales en la que está muy expresada (por ejemplo, la línea de células de carcinoma de colon cultivada, T84) utilizando métodos establecidos [J. E. Langridge, et al, *Biochim. Biophys. Acts.* 751: 318 (1983)]. Las proteínas periféricas de estas membranas son las que se separan por extracción con altas concentraciones salinas, un elevado pH o agentes caótrofos tales como diyodosalicilato de litio. La totalidad de las proteínas integrales que permanecen, incluida la proteína CFTR, se solubilizan utilizando un detergente tal como octil-glucósido [Landry, et al, *supra*], CHAPS [D. J. Beros et al, *J. Biol. Chem.* 262:10613 (1987)] u otros compuestos de acción similar. Haciendo uso de los dominios de unión de nucleótidos de CFTR, se utiliza luego la cromatografía de afinidad con azul cibacron [S. T. Thompson et al. *Proc. Natl. Acad. Sci. USA* 72: 669 (1975)] para unir la proteína CFTR y separarla de otras proteínas integrales de la mezcla de detergentes estabilizada. Dado que CFTR es una glicoproteína, la cromatografía con lectina diferencial puede lograr una purificación adicional [Riordan et al. *J. Biol. Chem.* 254:1270 (1979)]. La purificación final hasta la homogeneidad se consigue luego utilizando otros procesos de purificación de proteínas convencionales; es decir, la cromatografía de intercambio de iones, la cromatografía de permeación en gel, cromatografía de adsorción o enfoque isoelectrico, según sea necesario. Alternativamente, se hace uso de procesos de purificación de una sola etapa tales como cromatografía de inmuno-afinidad, utilizando anticuerpos inmovilizados contra la proteína CFTR (o fragmentos de la misma) o electroforesis en gel de poliácridamida preparativa utilizando una instrumentación avanzada tal como el "230A HPEC System" de Applied Biosystems. Basado en la experiencia en la purificación de la glicoproteína P [Riordan et al, *supra*], otro miembro de la categoría general de

las proteínas de la membrana asociadas al transporte de unión a nucleótidos, se facilita la purificación de la proteína CFTR.

Además de la purificación a partir de tejidos y células en los que la proteína CFTR está muy expresada, se utilizan procedimientos similares para purificar CFTR a partir de células transfectadas con vectores que contienen el gen de CF (ADNc) según se describe antes. Productos proteínicos que resultan de la expresión de la versión modificada de la secuencia de ADNc se purifican de una manera similar. Los criterios de la homogeneidad de proteínas, así proporcionados, incluyen los convencionales para el campo de la química de proteínas, incluida la electroforesis en gel monodimensional y bidimensional y la terminación de aminoácidos N-terminal. La proteína purificada se utiliza en análisis físicos bioquímicos adicionales para determinar rasgos de su estructura secundaria y terciaria, para ayudar al diseño de fármacos, para fomentar el funcionamiento adecuado de las formas de CF mutantes. En la preparación para uso en la terapia de proteínas, se considera la ausencia de sustancias contaminantes potencialmente tóxicas. Se reconoce que la naturaleza hidrófoba de la proteína necesita la inclusión de compuestos anfífilos tales como detergentes y otros [J. V. Ambud Kar y P. C. Maloney *J. Biol. Chem.* 261:10079 (1986)] en todas las fases de su manipulación.

5.0 Rastreo de CF

5.1 Diagnósis basada en ADN

Dado el conocimiento de la mutación principal según se describe en esta memoria, el rastreo de portadores y la diagnósis prenatal se pueden llevar a cabo como sigue.

La población de alto riesgo de una fibrosis quística es la caucasiana. Por ejemplo, cada mujer y/u hombre caucasiano de edad progeneradora sería rastreado con el fin de determinar si ella o él era un portador (aproximadamente un 5% de probabilidad para cada individuo). Si ambos son portadores, son entonces una pareja de riesgo para un niño con fibrosis quística. Cada niño de la pareja de riesgo tiene un 25% de posibilidad de verse afectado por una fibrosis quística. El proceso para determinar el estado portador utilizando las sondas descritas en esta memoria es como sigue.

Una aplicación principal de la información de la secuencia de ADN de los genes de CF normal y mutante se encuentra en el área del ensayo genético, detección del portador y la diagnósis prenatal. Individuos que portan mutaciones en el gen de CF (portador de la enfermedad o pacientes) se pueden detectar al nivel de ADN con el uso de una diversidad de técnicas. El ADN genómico utilizado para la diagnósis se puede obtener a partir de células del cuerpo tales como las que están presentes en la sangre periférica, orina, saliva, biopsia de tejidos, instrumental quirúrgico y material de autopsia. El ADN se puede utilizar directamente para la detección de la secuencia específica o se puede amplificar enzimáticamente *in vitro* utilizando PCR (Saiki et al. *Science* 230: 1350-1353, (1985), Saiki et al. *Nature* 324: 163-166 (1986)) antes del análisis. El ARN o su forma ADNc también se pueden utilizar para el mismo fin. Revisiones recientes de este objeto han sido presentadas por Caskey, (*Science* 236: 1223-8 (1989)) y por Landegren et al (*Science* 242: 229-237 (1989)).

La detección de la secuencia de ADN específica se puede lograr por métodos tales como la hibridación utilizando oligonucleótidos específicos (Wallace et al. *Cold Spring Harbour Symp. Quant. Biol.* 51: 257-261 (1986)), la secuenciación de ADN directa (Church y Gilbert, *Proc. Natl. Acad. Sci. USA* 81: 1991-1995 (1988)), el uso de enzimas de restricción (Flavell et al. *Cell* 15: 25 (1978), Geever et al. *Proc. Natl. Acad. Sci. USA* 78: 5081 (1981)), la discriminación sobre la base de la movilidad electroforética en geles con reactivo desnaturante (Myers y Maniatis, *Cold Spring Harbour Symp. Quant. Biol.* 51: 275-284 (1986)), la protección de RNasa (Myers, R. M., Larin, J. y T. Maniatis *Science* 230: 1242 (1985)), la escisión química (Cotton et al. *Proc. Nat. Acad. Sci. USA* 85: 4397-4401 (1985)) y el proceso de detección mediada por ligasa [Landegren et al. *Science* 241: 1077 (1988)].

Oligonucleótidos específicos para las secuencias normal o mutante se sintetizan químicamente utilizando máquinas comercialmente disponibles, marcadas radiactivamente con isótopos (tales como ³²P) o no radiactivamente (con marcadores tales como biotina (Ward y Langer et al. *Proc. Nat. Acad. Sci. USA* 78: 6633-6657 (1981)) y se hibridan a muestras de ADN individuales inmovilizadas sobre membranas u otros soportes sólidos mediante punto-borrón o transferencia a partir de geles después de la electroforesis. La presencia o ausencia de estas secuencias específicas se visualiza por métodos tales como autorradiografía o reacciones fluorométricas (Landegren et al, 1989, *supra*) o colorimétricas (Gebeyehu et al. *Nucleic Acids Research* 15: 4513-4534 (1987)). Una realización de este método de rastreo de oligonucleótidos ha sido aplicada en la detección de la delección de F508 según se describe en esta memoria.

Las diferencias de secuencia entre individuos normales y mutantes se puede revelar mediante el método de secuenciación de ADN directa de Church y Gilbert (*supra*). Segmentos de ADN clonados se pueden utilizar como sondas para detectar segmentos de ADN específicos. La sensibilidad de este método se ve muy reforzada cuando se combina con PCR [Wrichnik et al. *Nucleic Acids Res.* 15: 1529-542 (1987); Wong et al, *Nature* 330:384-386 (1987); Stofflet et al, *Science* 239: 491-494 (1988)]. En este último procedimiento, se utiliza un cebador de la secuenciación que se encuentra dentro de la secuencia amplificada con un producto de PCR de doble cadena o una plantilla de cadena sencilla generada por una PCR modificada. La determinación de la secuencia se efectúa por procesos convencionales con nucleótidos radiomarcados o mediante procesos de secuenciación automática con marcadores fluorescentes.

Las alteraciones de la secuencia pueden generar ocasionalmente sitios de reconocimiento de enzimas de restricción fortuitos que se revelan mediante el uso de una digestión con enzimas apropiada, seguido de una hibridación de

borrones en gel convencional (Southern, J. Mol. Biol. 98: 503 (1975)). Fragmentos de ADN que portan el sitio (ya sea normal o mutante) se detectan por reducción de tamaño o aumento en el correspondiente número de fragmentos de restricción. Muestras de ADN genómico también se pueden amplificar por PCR antes del tratamiento con la enzima de restricción apropiada; a continuación, se visualizan fragmentos de diferente tamaño bajo la luz UV en presencia de bromuro de etidio después de electroforesis en gel.

El ensayo genético basado en las diferencias de la secuencia de ADN se puede lograr por detección de la alteración en la movilidad electroforética de fragmentos de ADN en geles con o sin reactivo desnaturizante. Se pueden visualizar pequeñas deleciones e inserciones en la secuencia mediante electroforesis en gel de alta resolución. Por ejemplo, el producto de PCR con la deleción de 3 pb es claramente distinguible de la secuencia normal en un gel de poliacrilamida no desnaturizante al 8%. Los fragmentos de ADN de diferentes composiciones de secuencia se pueden distinguir sobre gel de gradiente de formamida desnaturizante en el que las movilidades de diferentes fragmentos de ADN se retardan en el gel a diferentes posiciones de acuerdo con sus temperaturas de "fusión parcial" específicas (Myers, *supra*). Además, las alteraciones de la secuencia, en particular pequeñas deleciones, se pueden detectar en forma de cambios en el modelo de migración de heteroduplex de ADN en una electroforesis en gel no desnaturizante, según se ha detectado para la mutación de 3 pb (F508) y en otros sistemas experimentales [Nagamine et al, Am. J. Hum. Genet. 45: 337-339 (1989)]. Alternativamente, un método para detectar una mutación que comprende una sustitución de una sola base u otro cambio pequeño podría basarse en la longitud del cebador diferencial en una PCR. Por ejemplo, se podría utilizar un cebador invariable además de un cebador específico para una mutación. Los productos de PCR de los genes normal y mutante se pueden luego detectar diferencialmente en geles de acrilamida.

Los cambios de la secuencia en lugares específicos también se pueden revelar mediante análisis de protección con nucleasa tal como RNasa (Myers, *supra*) y protección con S1 (Berk, A. J. y P. A. Sharpe Proc. nat. Acad. Sci. USA 75: 1274 (1978)), el método de escisión química (Cotton, *supra*) o el procedimiento de detección mediado por ligasa (Landegren, *supra*).

Además de una electroforesis en gel convencional y de los métodos de hibridación por transferencia, los fragmentos de ADN también se pueden visualizar por métodos en los que las muestras de ADN individuales no están inmovilizadas sobre membranas. Las secuencias de la sonda y diana pueden estar ambas en solución, o la secuencia de la sonda puede estar inmovilizada [Saiki et al, Proc. Natl. Acad. Sci. USA, 86:6230-6234 (1989)]. Para identificar genotipos individuales específicos se puede utilizar una diversidad de métodos de detección tales como autorradiografía con intervención de radioisótopos, detección directa de desintegración radiactiva (en presencia o ausencia de agente de centelleo), espectrofotometría que implica reacciones colorígenas y fluorometría que implica reacciones fluorógenas.

Dado que en el gen de CF se anticipa más de una mutación, un sistema múltiple es un protocolo ideal para el rastreo de portadores de CF en la detección de mutaciones específicas. Por ejemplo, se puede utilizar una PCR con múltiples cebadores de oligonucleótidos específicos y sondas de hibridación para identificar todas las posibles mutaciones al mismo tiempo (Chamberlain et al. Nucleic Acids Research 16: 1141-1155 (1988)). El proceso puede implicar sondas de oligonucleótidos específicas para la secuencia inmovilizadas (Saiki et al, *supra*).

5.2 Detección de la mutación principal

Estos métodos de detección se pueden aplicar a la diagnosis prenatal utilizando células del fluido amniótico, la biopsia de los vellos coriónicos o la clasificación de células fetales procedentes de la circulación materna. El ensayo de portadores de CF de población se puede incorporar como un componente esencial en un programa de ensayo genético a gran escala para enfermedades comunes.

De acuerdo con una realización de la invención, la porción del segmento de ADN que es informativa de una mutación, tal como una mutación de acuerdo con esta realización, es decir la porción que rodea inmediatamente a la deleción F508, se puede luego amplificar utilizando técnicas de PCR convencionales [según la revisión en Landegren, Ulf, Robert Kaiser, C. Thomas Caskey y Leroy Hood, DNA Diagnostics-Molecular Techniques and Automation, en Science 242: 229-237 (1988)]. Se contempla que la porción del segmento de ADN que se utiliza pueda ser un segmento de ADN sencillo o una mezcla de diferentes segmentos de ADN. Sigue ahora una descripción detallada de esta técnica.

Se ha de rastrear una región específica de ADN genómico procedente de la persona o feto. Una región específica de este tipo se define por los cebadores de oligonucleótidos C16B (5'GTTTTCTGGATTATGCCTGGGCAC3') y C16D (5'GTTGGCATGCTTTGATGACGCTTC3'). Las regiones específicas se amplificaron mediante la reacción en cadena de la polimerasa (PCR). 200-400 ng de ADN genómico, procedente de linfoblastos cultivados o muestras de sangre periférica de individuos CF y sus progenitores se utilizaron en cada PCR con los cebadores de oligonucleótidos indicados anteriormente. Los oligonucleótidos se purificaron con Oligonucleotide Purification Cartridges® (Applied Biosystems) o columnas NENSORB® PREP (Dupont) con procedimientos recomendados por los suministradores. Los cebadores se reasociaron a 62°C durante 45 s, se extendieron a 72°C durante 120 s (con 2 unidades de Taq ADN polimerasa) y se desnaturizaron a 94°C durante 60 s durante 28 ciclos con un ciclo final de 7 min para la extensión en un termociclador automático Perkin-Elmer/Cetus con un programa Step-Cycle (ajustes de transición a 1,5 min). Porciones de los productos de PCR se separaron mediante electroforesis en geles de agarosa al 1,4%, se transfirieron a una membrana Zetabind® (Biorad) de acuerdo con procesos convencionales. Las dos sondas de oligonucleótidos de la figura 15 (en cada caso 10 ng) se marcaron por separado con 10 unidades de T4 polinucleótido quinasa (Pharmacia) en una reacción de 10 µl que contenía Tris-HCl 50 mM (pH 7,6), MgCl₂ 10 mM, ditiotreitól 0,5 mM, espermidina

ES 2 210 224 T3

10 mM, EDTA 1 mM y 30-40 μCi de γ [^{32}P] - ATP durante 20-30 min a 37°C. Los radionucleótidos no incorporados se separaron con una columna Sephadex G-25 antes del uso. Las condiciones de hibridación eran como las descritas previamente (J. M. Rommens et al *Am. J. Hum. Genet.* 43,645 (1988)), excepto que la temperatura era de 37°C. Las membranas se lavaron dos veces a la temperatura ambiente con 5 x SSC y dos veces a 39°C con 2 x SSC (1 x SSC = NaCl 150 mM y citrato de Na 15 mM). La autorradiografía se efectuó a la temperatura ambiente durante una noche. Las autorradiografías muestran los resultados de hibridación de ADN genómico con las 2 sondas de oligonucleótidos específicas según se indica en la figura 15. La sonda C detecta la secuencia de ADN normal y la sonda F detecta la secuencia mutante. Una muestra de ADN genómico procedente de cada miembro de familia se amplificó mediante la reacción en cadena de la polimerasa, y los productos se separaron por electroforesis sobre un gel de agarosa al 1,4% y luego se transfirieron a una membrana Zetabind (Biorad) de acuerdo con procesos convencionales. El blanco de agua y el ADN del plásmido, T16 y C16, correspondientes a la secuencia normal (N) y a la delección F508 (CF), respectivamente, se incluyeron como testigos.

La delección de 3 pb también fue revelada mediante electroforesis en gel de poliacrilamida. Cuando la PCR generada por los cebadores C16B y C16D antes mencionados se aplicó a un gel de poliacrilamida al 8%, se electroforizó durante 2 h a 20 V/cm en un tampón Tris-borato 90 mM (pH 8,3) fragmentos de ADN de una movilidad diferente eran claramente detectables para individuos sin la delección de 3 pb, heterocigóticos u homocigóticos para la delección. Además, en los heterocigóticos se observó una banda de ADN extra, presumiblemente el heterodúplex entre hebras de ADN normal y mutante. Una alteración similar en la movilidad del gel para heterodúplex formados durante PCR también ha sido reseñada para sistemas experimentales en los que están implicadas pequeñas delecciones (Nagamine et al *supra*). Estos desplazamientos de la movilidad se pueden utilizar como base para los ensayos de rastreo genéticos radiactivos.

5.3 Programas de rastreo de CF

Se aprecia que únicamente el 70% de los portadores puede detectarse utilizando las sondas de F508 específicas de esta realización particular de la invención. Así, si un individuo ensayado no es un portador utilizando las sondas F508, su estado portador no puede ser excluido, pueden portar alguna otra mutación como se ha señalado previamente. Sin embargo, si tanto el individuo como el cónyuge del individuo sometido a ensayo son portadores de la mutación F508, se puede establecer con certeza que se trata de una pareja de riesgo. La secuencia del gen según se describe en esta memoria es un pre-requisito esencial para la determinación de las otras mutaciones.

La diagnosis prenatal es una extensión lógica del rastreo de portadores. Una pareja puede ser identificada como de riesgo por tener un niño con fibrosis quística en una de dos maneras: si ya tienen un niño con fibrosis quística, ambos son, por definición, portadores obligados de la enfermedad, y cada niño subsiguiente tiene un 25% de posibilidades de verse afectado por la fibrosis quística. Una ventaja principal de la presente invención elimina la necesidad del análisis del pedigrí de la familia, mientras que de acuerdo con esta invención, un programa de rastreo por mutación de genes como se ha esbozado anteriormente u otro método similar se pueden utilizar para identificar una mutación genética que conduzca a una proteína con función alterada. Esto no depende de una averiguación previa de la familia a través del niño afectado. Por ejemplo, muestras de ADN fetales se pueden obtener, como se ha mencionado previamente, a partir de células de fluido amniótico y muestras de vello coriónico. La amplificación por técnicas de PCR convencionales puede entonces realizarse en este ADN de la plantilla.

Si se demuestra que los dos padres son portadores con la delección F508, la interpretación de los resultados sería la siguiente. Si existe una hibridación del ADN fetal a la sonda normal (sin delección, como se muestra en la figura 15), el feto no se verá afectado por la fibrosis quística, a pesar de que puede ser un portador de CF (50% de probabilidades para cada feto de una pareja de riesgo). Si el ADN fetal se hibrida sólo a la sonda de delección F508 y no a la sonda normal (como se muestra en la figura 15), el feto se verá afectado por una fibrosis quística.

Se aprecia que para esta y otras mutaciones en el gen de CF, se puede utilizar una gama de procesos específicos diferentes para proporcionar una diagnosis completa para todos los portadores de CF o pacientes CF potenciales. Una descripción completa de estos procedimientos se describe más adelante.

Por lo tanto, la invención proporciona un método y un estuche para determinar si un sujeto es un portador de CF o un paciente CF. En resumen, el método de rastreo comprende las etapas de:

proporcionar una muestra biológica del sujeto a rastrear; y proporcionar un análisis para detectar en la muestra biológica la presencia de al menos un miembro del grupo que consiste en el gen de CF normal, productos génicos de CF normales, un gen de CF mutante, productos génicos de CF mutantes y sus mezclas.

El método se puede caracterizar, además, por incluir al menos una o más sondas de nucleótidos que es un fragmento de secuencia de ADN diferente, por ejemplo, el ADN de la figura 1, o un fragmento de la secuencia de ADN diferente del cromosoma 7 humano y que está situado en cualquier lado de la secuencia de ADN de la figura 1.

Un estuche, de acuerdo con una realización de la invención, adecuado para uso en la técnica de rastreo para analizar la presencia del gen de CF por un inmunoensayo comprende:

- (a) un anticuerpo que se une específicamente a un producto génico del gen de CF;

ES 2 210 224 T3

- (b) medios de reactivos para detectar la unión del anticuerpo al producto génico; y
- (c) estando presentes el anticuerpo y los medios de reactivos en cantidades eficaces para efectuar el inmunoensayo.

5 El estuche para analizar la presencia del gen de CF también se puede proporcionar mediante técnicas de hibridación. El estuche comprende:

- (a) una sonda de oligonucleótidos que específicamente se une al gen de CF;
- 10 (b) medios de reactivos para detectar la hibridación de la sonda de oligonucleótidos al gen de CF; y
- (c) estando presente la sonda y los medios reactivos en cada caso en cantidades eficaces para llevar a cabo el análisis de hibridación.

15 5.4 Anticuerpos para detectar CFTR

Como se ha mencionado, se desarrollan anticuerpos contra epítomos dentro de la proteína CFTR para proporcionar una información amplia de las características de la proteína y otra información valiosa, que incluye:

- 20 1. Para permitir la visualización de la proteína en células y tejidos en los que se expresa mediante inmunotransferencia ("borrones de Western") después de electroforesis en gel de poliacrilamida. Esto permite una estimación del tamaño molecular de la proteína madura incluida la contribución procedente de las células de restos añadidos después de la traducción y que incluyen cadenas de oligosacáridos y grupos fosfato, por ejemplo. Se pueden utilizar técnicas inmunocitoquímicas que incluyen inmunofluorescencia e inmuno-microscopía electrónica para establecer la localización subcelular de la proteína en las membranas celulares. Los anticuerpos también se pueden utilizar para proporcionar otra técnica para detectar cualesquiera de las otras mutaciones de CF que dan como resultado la síntesis de una proteína con un tamaño alterado.
- 25 2. Se pueden utilizar anticuerpos contra distintos dominios de la proteína para determinar la disposición topológica de la proteína en la membrana celular. Esto proporciona información sobre segmentos de la proteína que son accesibles a agentes moduladores externamente añadidos para fines de terapia con fármacos.
- 30 3. Las relaciones de estructura-función de porciones de la proteína se pueden examinar utilizando anticuerpos específicos. Por ejemplo, es posible introducir en células anticuerpos que reconocen cada uno de los bucles citoplásmicos cargados que unen las secuencias de la transmembrana, así como porciones de los pliegues de unión de nucleótidos y el dominio R. La influencia de estos anticuerpos sobre parámetros funcionales de la proteína proporciona una perspectiva en los mecanismos reguladores de las células y sugieren en potencia medios para modular la actividad de la proteína defectuosa en pacientes de CF.
- 35 4. Anticuerpos con la avidéz apropiada también permiten la inmunoprecipitación y la purificación por inmunofinidad de la proteína. La inmunoprecipitación facilitará la caracterización de la síntesis y la modificación post-traducción que incluye la unión de ATP y la fosforilación. Se requerirá la purificación para estudios de la estructura de la proteína y para la reconstitución de su función, así como la terapia basada en proteínas.
- 40

45 Con el fin de preparar los anticuerpos, se han sintetizado proteína de fusión que contienen porciones definidas de polipéptidos CFTR en bacterias mediante la expresión de correspondientes secuencias de ADN en un vehículo de clonación adecuado, mientras que péptidos más pequeños se sintetizaron químicamente según se describe en la Tabla 8. Las proteínas de fusión se purificaron, por ejemplo, mediante cromatografía de afinidad sobre glutation-agarosa, y los péptidos se acoplaron a una proteína portadora (hemocianina), se mezclaron con adyuvante de Freund y se inyectaron en conejos. Después de inyecciones de refuerzo a intervalos bisemanales, los conejos se desangraron y se aislaron los sueros. Las proteínas de fusión teñidas se muestran en las figuras 19a. La pista 1, plásmido control no inducido; la pista 2, plásmido control inducido por IPTG que expresa sólo glutation-S-transferasa (GST); la pista 3, banda de GST purificada por afinidad a 27 kilodalton (kD); la pista 4 está no inducida; la pista 5 está inducida y la pista 6 es la proteína de fusión purificada nº 1 de la Tabla 8. En la figura 19b, la electroforesis en gel es de lisados procedentes de bacterias transformadas con plásmidos pGEX que contienen proteínas de fusión nº 5 de la Tabla 8 para las pistas 1 y 2 y proteínas de fusión nº 2 de la Tabla 8 para las pistas 3 y 4. La pista 1 de la figura 19b es para el plásmido no inducido, mientras que la pista 2 es para que el plásmido inducido exprese la proteína de fusión nº 5. La pista 3 de la figura 19b es para el plásmido no inducido, mientras que la pista 4 es para que el plásmido inducido exprese la proteína de fusión nº 2. Los inmuno-borrones de la proteína de fusión nº 1 sondados con antisueros obtenidos de las segundas sangrías de dos conejos diferentes se muestran en la figura 20. La tinción es con el segundo anticuerpo conjugado con fosfatasa alcalina [Blake et al, Anal. Biochem. 136:175, (1984)]. Estos dos sueros inmunes tiñen la proteína de fusión de 32 kD, mientras que los sueros preinmunes no lo hacen. La figura 21 muestra la reactividad de uno de estos sueros inmunes con una banda de aproximadamente 200 kD de tamaño en membranas aisladas de células de carcinoma de colon T-84 que expresan el transcrito de CFTR a un elevado nivel. Esta banda está en el intervalo de tamaños que podría esperarse para la proteína CFTR que tiene un peso molecular predicho de 169 kD antes de las modificaciones posteriores a la traducción.

Sueros procedentes de conejos inmunizados con el conjugado LKH del péptido nº 2 se rastrearon de nuevo tanto

ES 2 210 224 T3

en cuanto a péptido puro como a KLH, como se muestra en la figura 22. En esta figura, H significa hemocianina; P1, péptido n° 1; P2, péptido n° 2. Se indican cantidades de proteína o péptido representadas en ng. Este antisuero detecta una cantidad tan pequeña como 1 ng del péptido y no reacciona en absoluto con el péptido testigo n° 1.

5 Así, es posible desarrollar anticuerpos policlonales específicos para las dos proteínas de fusión que contienen porciones de la proteína CFTR y péptidos que corresponden a segmentos cortos de su secuencia. De manera similar, se puede inyectar a ratones conjugados de KLH de los péptidos 1, 2 y 7 de la Tabla 8 para iniciar la producción de anticuerpos monoclonales contra estos segmentos de proteína CFTR. De manera similar, los anticuerpos monoclonales se pueden desarrollar contra otros dominios de la proteína CFTR.

10 En cuanto a la generación de anticuerpos policlonales, los inmunógenos para el desarrollo de anticuerpos monoclonales (amcs) contra la proteína CFTR son proteínas de fusión bacterianas [Smith et al, *Gene* 67:31 (1988)] que contienen porciones del polipéptido CFTR o péptidos sintéticos correspondientes a los segmentos cortos (12 a 25 aminoácidos de longitud) de la secuencia. La metodología esencial es de Kohler y Milstein [*Nature* 256: 495 (1975)].

15 Ratones Balb/c se inmunizan mediante inyección intraperitoneal con 500 μ g de proteína de fusión pura o péptidos sintéticos en adyuvante de Freund incompleto. Se administra una segunda inyección después de 14 días, una tercera al cabo de 21 días y una cuarta después de 28 días. Los animales individuales, así inmunizados, se sacrifican, una, dos y cuatro semanas después de la inyección final. Los bazo se separan, sus células se disocian, se recogen y se fusionan con células de mieloma Sp2/O-Ag14 de acuerdo con Geffer et al, *Somatic Cell Genetics* 3:231 (1977). La mezcla de fusión se distribuye en un medio de cultivo selectivo para la propagación de células fusionadas que se desarrollan hasta que tengan una confluencia de aproximadamente el 25%. En este instante, los sobrenadantes del cultivo se someten a ensayo en cuanto a la presencia de anticuerpos que reaccionan con un antígeno CFTR particular. A continuación, se usa luego un segundo anticuerpo anti-ratón marcado con fosfatasa alcalina para la detección de anticuerpos positivos. Después, las células procedentes de pocillos de cultivo positivos se expanden en el cultivo, sus sobrenadantes se recogen para el ensayo ulterior y las células se almacenan profundamente congeladas en medio con contenido en un agente crioprotector. Para obtener grandes cantidades de un amc, células productoras se inyectan en el peritoneo a razón de 5×10^6 células por animal, y se obtiene el fluido ascites. La purificación se realiza mediante cromatografía sobre proteína G- o proteína A-agarosa de acuerdo con Ey et al, *Immunochemistry* 15:429 (1977).

30 La reactividad de estos amcs con la proteína CFTR se confirma mediante electroforesis en gel de poliacrilamida de membranas aisladas a partir de células epiteliales en las que se expresa y mediante inmuno-transferencia [Towbin et al, *Proc. Natl. Acad. Sci. USA* 76: 4350 (1979)].

35 Además del uso de anticuerpos monoclonales específicos para cada uno de los diferentes dominios de la proteína CFTR para sondear sus funciones individuales, se utilizan otros amcs que pueden distinguir entre las formas normal y mutante de la proteína CFTR para detectar la proteína mutante en muestras de células epiteliales obtenidas de pacientes tales como "cepillos" de la biopsia de la mucosa nasal [R. De-Lough y J. Rutland, *J. Clin. Pathol.* 42, 613 (1989)] o glándulas sudoríparas que contienen muestras de la biopsia de la piel.

40 Anticuerpos capaces de esta distinción se obtienen mediante hibridomas de rastreo diferencial a partir de conjuntos emparejados de ratones inmunizados con un péptido que contiene la fenilalanina en la posición 508 del aminoácido (por ejemplo GTIKENIIFGVSY) o un péptido que es idéntico, excepto por la ausencia de F508 (GTIKENIIGVSY). Amcs capaces de reconocer las otras formas mutantes de la proteína CFTR presentes en pacientes además de o en lugar de la delección F508 se obtienen utilizando estrategias de producción de anticuerpos monoclonales similares.

45 Anticuerpos contra las versiones normal y CF de la proteína CFTR y de sus segmentos se utilizan en la microscopía óptica y la microscopía inmunoelectrónica diagnósticamente inmunocitoquímica e inmunofluorescente para demostrar la distribución de CFTR del tejido, celular y subcelular dentro de los órganos de pacientes y portadores de CF e individuos no CF.

50 Se utilizan anticuerpos para la modulación terapéutica al fomentar la actividad de la proteína CFTR en pacientes CF y en células de pacientes CF. Modos posibles de una modulación de este tipo podrían implicar la estimulación debida a la reticulación de moléculas de proteína CFTR con anticuerpos multivalentes en analogía con la estimulación de algunos receptores de la membrana de la superficie celular tales como el receptor de insulina [O'Brien et al, *Euro. Mol. Biol. Organ. J.* 6:4003 (1987)], el receptor del factor de crecimiento epidermal [Schreiber et al, *J. Biol. Chem.* 258: 846 (1983)] y moléculas asociadas al receptor de células T tal como CD4 [Veillette et al *Nature*, 338:257 (1989)].

60 Se utilizan anticuerpos para dirigir el suministro de agentes terapéuticos a las células que expresan una proteína CFTR defectuosa en CF. Para este fin, los anticuerpos se incorporan en un vehículo tal como un liposoma [Matthay et al, *Cancer Res.* 46:4904 (1986)] que porta el agente terapéutico tal como un fármaco o el gen normal.

ES 2 210 224 T3

TABLA 8

Fragmentos de CFTR utilizados para generar anticuerpos

5	Proteínas de fusión GST ^a que contienen residuos CFTR	Dominio de CFTR de la fig. 13
	1. 204-249	TM3, Ext. 2, TMA
	2. 347-698	NBF-1, 1/2 dominio R N-term.
10	3. 710-757	Centro del dominio R cargado neg.
	4. 758-796	Segmento del dominio R cargado pos.
	5. 1188-1480	Dominio cito. C-term. con NBF-2
15	Conjugados de KLH ^b que contienen péptidos CFTR:	
	1. 28-45	citoplásmico N-term.
	2. 58-75	citoplásmico N-term.
20	3. 104-117	1 ^{er} extracelular
	4. 139-153	2 ^o citoplásmico
	5. 279-294	N-term. de 3 ^{er} citoplásmico
	6. 500-512	NBF-1; en torno a la delección F508
25	7. 725-739	Centro cargado del dominio R
	8. 933-946	5 ^o citoplásmico
	9. 1066-1084	6 ^o citoplásmico

30 a. fragmentos de restricción que codifican estos fragmentos ligados al extremo 3' de glutatión S-transferasa (GST) de *Schistosoma japonicum* en el vector de expresión del plásmido pGEX según se identifica en Smith et al, Gene 67:31, (1988).

35 b. Péptidos acoplados a través de una cisteína N-terminal a la proteína portadora de hemocianina de lamprea bocallave (KLH-keyhole limpet hemocyanin) de acuerdo con Green et al Cell 28:477 (1982). TM significa secuencias de la transmembrana.

5.5 Análisis de RFLP

40 Esta invención proporciona un cierto número de beneficios que proceden directamente del descubrimiento y la caracterización del gen de CF que son de aplicación práctica inmediata. La secuencia de aminoácidos de CFTR proporciona una perspectiva en la estructura y la función de la proteína, así como los mecanismos moleculares en los que CFTR participa y que son defectuosos en la fibrosis quística. Esta información permite la generación de herramientas y conceptos adicionales en la investigación y la terapia de esta enfermedad.

45 La detección del portador, la diagnosis de ADN y la deliberación de la familia son algunas de las aplicaciones de la invención. El ensayo genético basado en ADN previo para CF ha estado principalmente disponible para familias con niños afectados y a sus parientes más próximos. El conocimiento de las mutaciones de CF al nivel de la secuencia de ADN permite el ensayo de cualquier individuo aleatorio; la estimación de los autores de la invención muestra que el 46% de los pacientes de CF sin una historia familiar previa pueden ser diagnosticados con precisión mediante análisis por ADN, y el 68% de los portadores de CF en la población puede ser identificado a través de la delección F508.

50 Dado que la frecuencia de portadores en la población norteamericana es aproximadamente 1 de 20, es factible rastrear a la totalidad de mujeres y/u hombres de edad progenitora, por ejemplo, en cuanto a su estado portador. La detección del portador utilizando sondas específicas para la delección F508 recogerá el 70% de los portadores. Los portadores restantes serán detectados mediante una batería de sondas específicas para los diversos grupos de haplotipo identificados anteriormente.

60 Dado que la delección F508 constituye aproximadamente el 70% de todas las mutaciones CF, se pueden utilizar análisis de RFLP suplementarios al ensayo de delección directa en cuanto a miembros de la familia o parientes próximos de pacientes de CF. Se espera que aproximadamente el 55% de los progenitores de CF que no porten la mutación F508 sea informativo para el marcador de ADN JG2E1 (KM19) [Kerem et al Am. J. Hum. Genet 44:827-834 (1989); Estivill et al, Genomics 1:257 (1987)] basado en un análisis retrospectivo de nuestras familias de enlace a CF; un 39% adicional sería informativo si también se ensayaran E6 (Taq I) [Kerem et al *supra*] y J3.11 (Msp I) [Wainright et al Nature (1985)]; virtualmente todos los progenitores serían informativos si se incluyeran H2.3 (XV2C-Taq I) [Kerem et al, *supra*; Estivill et al, Nature (1987)], E2.6 (E.9) (Msp I) [sonda disponible a petición], E4.1 (Mp6d.9) (Msp I) [sonda disponible a petición; Estivill et al, Am. J. Hum. Genet. (1989)], J44 (E3.1) (Xba I) [sonda disponible a petición] y metID (Ban I) [Spence et al, Am. J. Hum. Genet (1986), [ATCC n° 40219].

ES 2 210 224 T3

La utilidad de estas sondas radica en el hecho de que éstas se reconocen sitios de restricción polimórficos. Así, las sondas no se definen típicamente por su secuencia a lo largo del sitio polimórfico particular, sino más bien se pueden utilizar en base al conocimiento de las secuencias flanqueantes permitiendo la generación mediante reacción en cadena de polimerasa (PCR) de la región en cuestión, como sería conocido por un experto en la técnica.

Por ejemplo, la sonda E2.6 (Msp I) se define por completo por dos oligómeros flanqueantes: 5'GTGATCCAGTTT GCTCTCCA3' y 5'GGAATCACTCTTCCTGATAT3'. El uso de esta sonda E2.6 generada por PCR para detectar un polimorfismo en Msp I detectará dos alelos diferentes: ya sea un fragmento de 850 pb o un fragmento de 490 pb y un fragmento de 360 pb, dependiendo de la presencia o ausencia del sitio Msp I. De modo similar, la sonda J44 (E3.1) (Xba I) está definida por completo por dos oligómeros flanqueantes: 5'CAATGTGATTGGTGAAACTA3', y 5'CTTCTCCTCCTAGACACCTGCAT3'. El uso de esta sonda J44 (E3.1) generada por PCR para detectar un polimorfismo Xba I detectará dos alelos diferentes: ya sea un fragmento de 860 pb o un fragmento de 610 pb y un fragmento de 250 pb, dependiendo de la presencia o ausencia del sitio Xba I.

Los RFLPs enlazados también se pueden utilizar en el cálculo del riesgo para individuos que no portan la delección F508. En Beaudet et al Am. J. Hum. Genet 44:319-326 se ha comentado un procedimiento general de estimación de riesgo.

Para la diagnosis prenatal, el análisis de la enzima intestinal de los microvellos (Brock, Lancet 2: 941 (1983)) se puede realizar para aumentar la confianza de la diagnosis en los casos en los que la diagnosis de ADN es inconclusiva.

La diagnosis de ADN se utiliza actualmente para verificar si un feto nacerá con fibrosis quística, pero históricamente esto sólo se ha realizado después de que un grupo particular de padres haya tenido un niño con fibrosis quística y los identifique como portadores obligados. Sin embargo, en combinación con la detección del portador, según se ha esbozado antes, será posible la diagnosis de ADN para todos los partos de parejas portadoras. Si los progenitores ya han tenido un niño con fibrosis quística, se puede realizar un análisis del haplotipo extendido en el feto y, así, se reducirá grandemente el porcentaje de falso positivo o falso negativo. Si los padres no han tenido ya un niño afectado y la diagnosis de ADN en el feto se está llevando a cabo sobre la base de la detección del portador, se puede seguir realizando un análisis del haplotipo.

A pesar de que se ha pensado durante muchos años que existe una gran heterogeneidad clínica en la enfermedad fibrosis quística, ahora surgen dos categorías generales, denominadas suficiencia pancreática (CF-PS) e insuficiencia pancreática (CF-PI). Si las mutaciones relacionadas con estas categorías de enfermedad están bien caracterizadas, se puede asociar una mutación particular con un fenotipo clínico de la enfermedad. Esto permite cambios en el tratamiento de cada paciente. Así, la naturaleza de la mutación permitirá predecir, hasta una determinada magnitud, el pronóstico del paciente e indicar un tratamiento específico.

6.0 Biología molecular de la fibrosis quística

El postulado de que CFTR puede regular la actividad de los canales de iones, particularmente el canal Cl rectificador hacia afuera implicado como el defecto funcional en CF, se puede someter a ensayo mediante la inyección y traducción del ARNm de longitud completa de CFTR transcrito *in vitro* en oocitos de *Xenopus*. Los cambios que se producen en la corriente de iones a lo largo de la membrana de los oocitos se puede medir, dado que el potencial está establecido en un valor fijo. CFTR puede regular los canales endógenos de oocitos o puede ser necesario también introducir ARN de la célula epitelial para dirigir la traducción de las proteínas del canal. El uso de ARNm que codifica CFTR normal y CFTR mutante, según se proporciona mediante esta invención, hace posible estos experimentos.

Otros modos de expresión en el sistema de células heterólogo facilitan también la disección de las relaciones de estructura-función. La secuencia completa de ADN de CFTR ligada en un vector de expresión del plásmido se utiliza para transfectar células, de modo que se pueda verificar su influencia sobre el transporte de iones. Vectores de expresión del plásmido que contienen parte de la secuencia de CFTR normal junto con porciones de la secuencia modificada en sitios seleccionados se pueden utilizar en experimentos de mutagénesis *in vitro* realizados con el fin de identificar esas porciones de la proteína CFTR que son cruciales para la función reguladora.

6.1 Expresión de la secuencia de ADN

La secuencia de ADN se puede manipular en estudios para comprender la expresión del gen y su producto y, para conseguir la producción de grandes cantidades de la proteína para el análisis funcional, la producción de anticuerpos y la terapia de pacientes. Los cambios en la secuencia pueden o no alterar el modelo de expresión en términos de cantidades relativas, especificidad del tejido y propiedades funcionales. Las secuencias de ADNc parciales o de longitud completa, que codifican la proteína objeto, no modificada o modificada, se pueden limitar a vectores de expresión bacterianos tales como los plásmidos pRIT (Nilsson et al. EMBO J. 4:1075-1080 (1985)), pGEX (Smith y Johnson, Gene 67: 31-40 (1988)) o pATH (Spindler et al. J. Virol. 49: 132-141 (1984)) que se pueden introducir en células de *E. coli* para la producción de las correspondientes proteínas que se pueden aislar de acuerdo con los procedimientos de purificación de proteínas previamente comentados. La secuencia de ADN también se puede transferir desde su contexto existente a otros vehículos de clonación, tales como otros plásmidos, bacteriófagos, cósmidos, virus animales, cromosomas artificiales de levaduras (YAC- yeast artificial chromosomes) (Burke et al. Science 236: 806-812 (1987)), células somáticas y otros organismos simples o complejos tales como bacterias, hongos (Timberlake y

Marshall, Science 244: 1313-1317 (1989), invertebrados, plantas (Gasser y Fraley, Science 244: 1293 (1989) y cerdos (Pursel et al. Science 244: 1281-1288 (1989)).

Para la expresión en células de mamíferos, la secuencia de ADNc se puede ligar a promotores heterólogos tales como el virus de simio (SV) 40, promotor en el vector pSV2 [Mulligan y Berg, Proc. Natl. Acad. Sci USA, 78: 2072-2076 (1981)] y se pueden introducir en células tales como células de mono COS-1 [Gluzman, Cell, 23: 175-182 (1981)] para conseguir una expresión transitoria o a largo plazo. La integración estable de la construcción del gen quimérico puede mantenerse en células de mamífero mediante selección bioquímica tales como neomicina [Southern y Berg, J. Mol. Appln. Genet. 1: 327-341 (1982)] y ácido micofenólico [Mulligan y Berg, *supra*].

Las secuencias de ADN se pueden manipular con procedimientos convencionales tales como digestión con enzimas de restricción, relleno con ADN polimerasa, delección mediante exonucleasa, extensión mediante desoxinucleótido transferasa terminal, ligación de secuencias de ADN sintético o clonado, alteración de la secuencia dirigida al lugar a través de un intermedio bacteriófago de hebra sencilla o con el uso de oligonucleótidos específicos en combinación con PCR.

La secuencia de ADNc (o porciones derivadas de la misma) o un minigen (un ADNc con un intrón y su propio promotor) se introduce en vectores de expresión eucarióticos mediante técnicas convencionales. Estos vectores se diseñan para permitir la transcripción del ADNc en células eucarióticas y proporcionar secuencias reguladoras que inician y refuerzan la transcripción del ADNc y aseguran su adecuado corte y empalme y poliadenilación. Vectores que contienen las regiones de promotor y reforzador del virus de simio (SV) 40 o la repetición larga terminal (LTR-long terminal repeat) del virus del Sarcoma de Rous y la señal de poliadenilación y corte y empalme procedente de SV 40 están fácilmente disponibles [Mulligan et al Proc. Natl. Acad. Sci. USA 78: 1078-2076, (1981); Gorman et al Proc Natl. Acad. Sci USA 79: 6777-6781 (1982)]. Alternativamente, se puede utilizar el promotor endógeno de CFTR. El nivel de expresión del ADNc se puede manipular con este tipo de vector, ya sea utilizando promotores que tienen diferentes actividades (por ejemplo el baculovirus pAC373 puede expresar ADNcs a altos niveles en células de *S. frungiperda* [M. D. Summers y G. E. Smith en, Genetically Altered Viruses and the Environment (B. Fields, et al, compiladores) vol. 22 n° 319-328, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, Nueva York, 1985] o utilizando vectores que contienen promotores sujetos a modulación, por ejemplo el promotor que responde a glucocorticoides procedentes del virus del tumor mamario de ratón [Lee et al, Nature 294:228 (1982)]. La expresión del ADNc se puede vigilar en las células receptoras 24 a 72 horas después de la introducción (expresión transitoria).

Además, algunos vectores contienen marcadores seleccionables [tales como los genes bacterianos *gpt* [Mulligan et Berg *supra*] o *neo* [Southern y Berg J. Mol. Appln. Genet 1: 327-341 (1982)] que permiten el aislamiento de células mediante selección química, que tienen una expresión estable y a largo plazo de los vectores (y, por lo tanto, del ADNc) en la célula receptora. Los vectores se pueden mantener en las células en forma de entidades episomales y que replican libremente utilizando elementos reguladores de virus tales como papiloma [Sarver et al Mol. Cell Biol. 1:486 (1981)] o Epstein-Barr [Sugden et al Mol. Cell Biol. 5:410 (1985)]. Alternativamente, también se pueden producir líneas celulares que tienen integrado el vector en el ADN genómico. Estos dos tipos de líneas celulares proporcionan el producto génico sobre una base continua. También se pueden producir líneas celulares que tienen amplificado el número de copias del vector (y, por lo tanto, asimismo del ADNc) para crear líneas celulares que pueden producir altos niveles de producto génico [Alt et al. J. Biol. Chem. 253:1357 (1978)].

La transferencia de ADN en células eucarióticas, en particular humanas o de otros mamíferos es ahora una técnica convencional. Los vectores se introducen en las células receptoras en forma de ADN puro (transfección), por ejemplo por precipitación con fosfato de calcio [Graham y vander Eb, Virology 52:466 (1973) o fosfato de estroncio [Brash et al Mol. Cell Biol. 7 2013 (1987)], electroporación [Neumann et al EMBO J 1:841 (1982)], lipofección [Felgner et al Proc Natl. Acad. Sci USA 84:7413 (1987)], DEAE dextrano [McCuthan et al J. Natl Cancer Inst. 41:351 1968)], microinyección [Mueller et al Cell 15:579 (1978)], fusión de protoplastos [Schafner, Proc Natl. Aca. Sci USA 72: 2163] o escopetas de gránulos [Klein et al, Nature 327: 70 (1987)]. Alternativamente, el ADNc se puede introducir por infección con vectores de virus. Se desarrollan sistemas que utilizan, por ejemplo, retrovirus [Bernstein et al. Genetic Engineering 7: 235, (1985)], adenovirus [Ahmad et al J. Virol 57: 267 (1986)] o virus Herpes [Spaete et al Cell 30: 295 (1982)].

Estos sistemas de expresión eucarióticos se pueden utilizar para muchos estudios del gen de CF y el producto CFTR. Estos incluyen, por ejemplo: (1) la determinación de que el gen se expresa adecuadamente y que se han completado adecuadamente todas las modificaciones post-traducción necesarias para una actividad biológica completa, (2) identificar elementos reguladores situados en la región 5' del gen de CF y su papel en la regulación tisular o temporal de la expresión del gen de CF; (3) producción de grandes cantidades de la proteína normal para el aislamiento y la purificación; (4) para utilizar células que expresan la proteína CFTR en forma de un sistema de análisis para anticuerpos generados contra la proteína CFTR o un sistema de análisis para ensayar la eficacia de fármacos; (5) estudiar la función de la proteína completa normal, porciones específicas de la proteína o de proteínas mutantes que aparecen en la naturaleza o se producen artificialmente. Proteínas mutantes que se producen en la naturaleza existen en pacientes con CF, mientras que proteína mutante producida artificialmente puede diseñarse mediante alteraciones de la secuencia dirigida al lugar. Estos últimos estudios pueden demostrar la función de cualquier residuo aminoácido deseado mutando los nucleótidos que codifican ese aminoácido.

Al utilizar las técnicas anteriores, los vectores de expresión que contienen la secuencia del gen de CF o fragmentos

de la misma se pueden introducir en células humanas, células de mamíferos procedentes de otras especies o células de no mamíferos, según se desee. La elección de la célula se determina mediante el fin del tratamiento. Por ejemplo, se pueden utilizar células COS de mono [Gluzman, *Cell* 23:175 (1981)] que producen altos niveles del antígeno T de SV40 y permiten la replicación de vectores que contienen el origen de replicación de SV40, para demostrar que el vector puede expresar el producto proteínico, ya que no se requiere la función. Se podría realizar un tratamiento similar con fibroblastos de ovario de hamster chino (CHO) o 3T3 de NIH de ratón o con fibroblastos o linfoblastos humanos.

El vector de clonación recombinante de acuerdo con esta invención comprende entonces el ADN seleccionado de las secuencias de ADN de esta invención para la expresión en un huésped adecuado. El ADN está operativamente enlazado en el vector a una secuencia de control de la expresión en la molécula de ADN recombinante, de modo que se puede expresar el polipéptido de CFTR normal. La secuencia del control de la expresión se puede seleccionar del grupo que consiste en secuencias que controlan la expresión de genes de células procarióticas o eucarióticas y sus virus y combinaciones de los mismos. La secuencia para el control de la expresión se puede seleccionar específicamente del grupo que consiste en el sistema *lac*, el sistema *trp*, el sistema *tac*, el sistema *trc*, regiones de operador y promotor principales del fago lambda, la región control de la proteína de revestimiento fd, los promotores temprano y tardío de SV40, promotores derivados de polioma, adenovirus, retrovirus, baculovirus y virus de simio, el promotor para la 3-fosfoglicerato quinasa, los promotores de fosfatasa ácida de levadura, el promotor de los factores alfa-apareantes de levaduras y sus combinaciones.

La célula huésped, que se puede transfectar con el vector de esta invención, se puede seleccionar del grupo que consiste en *E. coli*, *Pseudomonas*, *Bacillus subtilis*, *Bacillus stearothermophilus* u otros bacilos; otras bacterias; levaduras; hongos; insectos; ratones u otros animales; o huéspedes vegetales; o células de tejidos humanos.

Se aprecia que para la secuencia de ADN mutante se emplean sistemas similares para expresar y producir el producto mutante.

6.2 Consideraciones de la función proteínica

Para estudiar la función de la proteína CFTR es preferible utilizar células epiteliales como receptores, ya que la expresión funcional adecuada puede requerir la presencia de otras vías o productos génicos que únicamente se expresan en este tipo de células. Células que se pueden utilizar incluyen, por ejemplo, líneas de células del epitelio humano tales como T84 (ATCC n° CRL 248) o PANC-1 (ATCC n° CLL 1469), o la línea de células del epitelio nasal de CF inmortalizada T43 [Jettan et al, *Science* (1989)] y células de pólipos o de las vías respiratorias nasales humanas primarias [Yanoshkes et al. *Ann. Rev. Resp. Dis.* 132: 1281 (1985)] o transformadas [Scholte et al. *Exp. Cell. Res.* 182: 559 (1989)], células pancreáticas [Harris y Coleman *J. Cell. Sci.* 87: 695 (1987)] o células de las glándulas sudoríparas [Collie et al. *In Vitro* 21: 597 (1985)] derivados de sujetos normales o de CF.

Las células de CF se pueden utilizar para someter a ensayo la actividad funcional de genes de CF mutantes. Análisis funcionales actuales disponibles incluyen el estudio del movimiento de aniones (Cl o I) a través de las membranas celulares en función de la estimulación de las células por agentes que elevan los niveles de AMP intracelular y activan los canales de cloruro [Stutto et al. *Proc. Nat. Acad. Sci. U.S.A.* 82: 6677 (1985)]. Otros análisis incluyen la medición de cambios en los potenciales celulares mediante pinzamientos zonales de la membrana de células enteras o de membranas aisladas [Frizzell et al. *Science* 233: 558 (1986), Welsch y Liedtke *Nature* 322: 467 (1986)] o el estudio de flujos de iones en láminas epiteliales de células confluentes [Widdicombe et al. *Proc. Nat. Acad. Sci.* 82: 6167 (1985)]. Alternativamente, el ARN preparado a partir del gen de CF podría inyectarse en oocitos de *Xenopus*. El oocito traducirá ARN en la proteína y permitirá su estudio. Ya que se desarrollan otros análisis más específicos, éstos también se pueden utilizar en el estudio de la función proteínica de CFTR transfectada.

Experimentos de “conmutación del dominio” entre CFTR y la glicoproteína P de resistencia a múltiples fármacos humana también se pueden efectuar para estudiar adicionalmente la proteína CFTR. En estos experimentos, vectores de expresión de plásmidos se construyen mediante técnicas rutinarias a partir de fragmentos de secuencia de CFTR y fragmentos de la secuencia de glicoproteína P ligados entre sí mediante ADN ligasa, de modo que se sintetizará una proteína que contiene las porciones respectivas de estas dos proteínas por parte de una célula huésped transfectada con el plásmido. Este último enfoque tiene la ventaja de que se pueden medir muchos parámetros experimentales asociados con la resistencia a múltiples fármacos. Por lo tanto, ahora es posible verificar la capacidad de segmentos de CFTR de influir en estos parámetros.

Estos estudios de influencia de CFTR sobre el transporte de iones servirá para llevar el campo del transporte epitelial al foso molecular. Esta es la primera molécula relacionada con el transporte de células epiteliales para la que se muestra la estructura primaria completa. El conocimiento de CFTR se puede utilizar para comprender mejor a un nivel molecular las características de la membrana de la célula epitelial en esta zona. Por ejemplo, las moléculas en la proximidad más estrecha a CFTR se pueden determinar mediante experimentos de reticulación. La hipótesis de que el papel de CFTR consiste en regular los canales de iones prediría que estos canales caerían necesariamente en esa categoría. Los grandes bancos de ADNc de alta calidad, construidos para la clonación de ADNcs de CFTR, también serán útiles para la clonación molecular de ADNcs para polipéptidos que constituyen otros sistemas de transporte de iones epiteliales, incluidos otros canales así como sistemas de co-transporte, contra-transporte y transporte activo.

6.3 Terapias

Se entiende que el objetivo principal de los diversos estudios bioquímicos utilizando las composiciones de esta invención es el desarrollo de terapias para evitar o solucionar el defecto de CF, utilizando tanto los enfoques farmacológicos como de “terapia génica”.

En el enfoque farmacológico, se buscan fármacos que eviten o superen el defecto de CF. Inicialmente, los compuestos se pueden someter a ensayo esencialmente al azar y se requieren sistemas de rastreo para discriminar entre muchos compuestos candidato. Esta invención proporciona sistemas de células huésped, que expresan varios de los genes de CF mutantes, que son particularmente bien adecuados para el uso como sistemas de rastreo de primer nivel. Preferiblemente, en el procedimiento de rastreo se utiliza un sistema de cultivo celular que utiliza células de mamífero (lo más preferiblemente células humanas) transfectadas con un vector de expresión que comprende una secuencia de ADN que codifica una proteína CFTR que contiene una mutación generadora de CF, por ejemplo la delección F508. Fármacos candidatos se someten a ensayo incubando las células en presencia del fármaco candidato y midiendo las funciones celulares dependientes de CFTR, especialmente midiendo corrientes de iones en las que el potencial de la membrana se establece a un valor fijo. Sin embargo, para alojar un gran número de análisis, análisis más convenientes se basan, por ejemplo, en el uso de colorantes fluorescentes sensibles a iones. Para detectar cambios en la concentración del ion Cl^- , son útiles SPQ o sus análogos.

Alternativamente, se podría utilizar un sistema exento de células. CFTR purificada podría reconstituirse en membranas artificiales y los fármacos podrían rastreadarse en un análisis exento de células [Al-Aqwatt, Science (1989)].

Al segundo nivel, se requiere el ensayo con animales. Es posible desarrollar un modelo de CF interfiriendo con la expresión normal de la parte antagonista del gen de CF en un animal tal como el ratón. El “noqueo” de este gen al introducir una forma mutante del mismo en la línea germinal de animales proporcionará una raza de animales con síndromes similares a CF. Esto permite el ensayo de fármacos que mostraron una expectativa en el rastreo basado en células del primer nivel.

Dado que se adquiere un conocimiento adicional sobre la naturaleza de la proteína y su función, será posible predecir estructuras de proteínas o de otros compuestos que interactúen con la proteína CFTR. Esto, a su vez, permitirá realizar determinadas predicciones sobre fármacos potenciales que interactuarán con esta proteína y que tienen un cierto efecto sobre el tratamiento de los pacientes. En última instancia, fármacos de este tipo pueden ser diseñados y sintetizados químicamente sobre la base de estructuras que se predice se requieren para interactuar con dominios de CFTR. Este enfoque se revisa en Capsey y Delvatte, Genetically Engineered Human Therapeutic Drugs Stockton Press, Nueva York, 1988. Estos fármacos potenciales también deben ser ensayados en el sistema de rastreo.

6.3.1 Terapia de reemplazamiento de proteínas

El tratamiento de CF se puede realizar reemplazando la proteína defectuosa por una proteína normal, o modulando la función de la proteína defectuosa o modificando otra etapa en la vía en la que participa CFTR con el fin de corregir la anomalía fisiológica.

Con el fin de poder reemplazar la proteína defectuosa por la versión normal, se ha de disponer de cantidades razonablemente grandes de proteína CFTR pura. La proteína pura se puede obtener según se describe anteriormente a partir de sistemas de células cultivadas. El suministro de la proteína al tejido de las vías respiratorias afectado requerirá su empaquetamiento en vesículas con contenido en lípidos que facilitan la incorporación de la proteína en la membrana celular. También puede ser factible utilizar vehículos que incorporen proteínas tales como proteína tensioactiva, tal como SAP(Val) o SAP(Phe) que efectúen esta función de modo natural, al menos para las células alveolares de los pulmones. (Solicitud de patente PCT WO/8803170, Whitsett et al, 7 de mayo de 1988 y solicitud de patente PCT WO 89/04327, Benson et al, 18 de mayo de 1989). Las vesículas con contenido en CFTR se introducen en las vías respiratorias por inhalación o irrigación, técnicas que actualmente se utilizan en el tratamiento de CF (Boat et al, *supra*).

6.3.2 Terapia con fármacos

La modulación de la función de CFTR se puede conseguir mediante el uso de agentes terapéuticos (fármacos). Éstos se pueden identificar mediante enfoques aleatorios utilizando un programa de rastreo en el que se vigile *in vitro* su eficacia en modular la proteína CFTR defectuosa. Los programas de rastreo pueden utilizar sistemas de células cultivadas en los que se expresa la proteína CFTR defectuosa. Alternativamente, se pueden diseñar fármacos para modular la actividad de CFTR conociendo las correlaciones de estructura y función de la proteína CFTR y conociendo el defecto específico en las diversas proteínas mutantes CFTR (Capsey y Delvatte, *supra*). Es posible que cada proteína CFTR mutante requiera un fármaco diferente para la modulación específica. Será entonces necesario identificar la o las mutaciones específicas en cada paciente de CF antes de iniciar la terapia con el fármaco.

Los fármacos se pueden diseñar para interactuar con diferentes aspectos de estructura o función de la proteína CFTR. Por ejemplo, un fármaco (o anticuerpo) puede unirse a un pliegue estructural de la proteína para corregir una estructura defectuosa. Alternativamente, un fármaco podría unirse a un residuo funcional específico y aumentar su

afinidad por un sustrato o cofactor. Dado que se conoce que miembros de la clase de proteínas con las que CFTR tiene una homología estructural pueden interactuar con, unirse a y transportar una diversidad de fármacos, es razonable esperar que terapias relacionadas con el fármaco puedan ser eficaces en el tratamiento de CF.

5 Un tercer mecanismo para reforzar la actividad de un fármaco eficaz sería modular la producción o la estabilidad de CFTR en el interior de la célula. Este aumento en la cantidad de CFTR podría compensar su función defectuosa.

La terapia con fármacos también se puede utilizar para compensar la función defectuosa de CFTR mediante interacciones con otros componentes de la vía fisiológica o bioquímica necesarios para la expresión de la función de CFTR. Estas interacciones pueden conducir a incrementos o disminuciones en la actividad de estas proteínas auxiliares. Los métodos para la identificación de estos fármacos serían similares a los descritos anteriormente para fármacos relacionados con CFTR.

En otros trastornos genéticos, ha sido posible corregir las consecuencias de funciones alteradas o normales carentes mediante el uso de modificaciones en la dieta. Esto ha adoptado la forma de separación de metabolitos, como es el caso de la fenilcetonuria, en la que la fenilalanina se elimina de la dieta en los primeros cinco años de vida para prevenir un retardo mental, o mediante la adición de grandes cantidades de metabolitos a la dieta, como es el caso de la deficiencia de adenosina desaminasa, en la que la corrección funcional de la actividad de la enzima se puede producir mediante la adición de enzima a la dieta. Así, una vez que se han elucidado los detalles de la función de CFTR y se ha definido el defecto básico en CF, la terapia se puede lograr mediante manipulaciones en la dieta.

El segundo enfoque terapéutico potencial es la denominada “terapia génica” en la que copias normales del gen de CF se introducen en pacientes con el fin de codificar con éxito la proteína normal en las células epiteliales claves de tejidos afectados. Lo más crucial es intentar lograr esto con las células epiteliales de las vías respiratorias del tracto respiratorio. El gen de CF se suministra a estas células en una forma en la que puede ser absorbido y codificar proteína suficiente para proporcionar una función reguladora. Como resultado, la calidad y tiempo de vida del paciente se verán grandemente ampliados. En última instancia, naturalmente, el objetivo es suministrar el gen a todos los tejidos afectados.

30 6.3.3 Terapia génica

Un enfoque para la terapia de CF es insertar una versión normal del gen de CF en el epitelio de las vías respiratorias de pacientes afectados. Es importante señalar que el sistema respiratorio es la causa primaria de morbilidad y mortalidad en CF; mientras que la enfermedad pancreática es un rasgo principal, éste es tratado actualmente relativamente bien con una suplementación enzimática. Así, la terapia génica con células somáticas [para una revisión, véase T. Friedmann, *Science* 244: 1275 (1989)] dirigida a las vías respiratorias aliviaría los problemas más graves asociados con la CF.

A. Vectores retrovirales. Los retrovirus han sido considerados el vector preferido para experimentos en la terapia génica somática, con una elevada eficacia de infección e integración y expresión estables [Orkin et al *Prog. Med. Genet* 7: 130 (1988)]. Un inconveniente posible es que es necesaria la división celular para la integración retroviral, de modo que puede ser que las células objetivadas en las vías respiratorias tengan que ser empujadas en el ciclo celular antes de la infección retroviral, quizás por medios químicos. El ADNc del gen de CF de longitud completa se puede clonar en un vector retroviral e impulsar desde su promotor endógeno o a partir de la LRT (long terminal repeat - repetición terminal larga) retroviral. Sería de esperar que la expresión de niveles de la proteína normal tan bajo como 10% de la proteína mutante endógena en pacientes de CF fuese beneficiosa, ya que ésta es una enfermedad recesiva. El suministro del virus puede conseguirse mediante aerosol o instilación en la tráquea.

B. Otros vectores virales. Otros sistemas de suministro que se pueden utilizar incluyen un virus asociado a adeno [AAV, McLaughlin et al, *J. Virol* 62: 1963 (1988)], virus de la vacuena [Moss et al *Annu. Rev. Immunol*, 5: 305, 1987)], papilomavirus bovino [Rasmussen et al, *Methods Enzymol* 139: 642 (1987)] o un miembro del grupo de los virus herpes tales como el virus Epstein-Barr (Margolske et al *Mol. Cell. Biol* 8:2937 (1988)]. A pesar de que se necesitaría aprender mucho sobre su biología básica, es posible la idea de utilizar un vector viral con un tropismo natural para las vías respiratorias (por ejemplo virus sincitial respiratorio, ecovirus, virus Cocksackie, etc.).

C. Transferencia génica no viral. También pueden ser productivos otros métodos de insertar el gen de CF en el epitelio respiratorio; muchos de estos son métodos de menor eficacia y requerirían en potencia una infección *in vitro*, una selección de los transfectantes y una reimplantación. Esto incluiría fosfato de calcio, DEAE dextrano, electroporación y fusión de protoplastos. Una idea particularmente atractiva es el uso de liposomas, que podría ser posible llevar a cabo *in vivo* [Ostro, *Liposomes*, Marcel-Dekker, 1987]. Lípidos catiónicos sintéticos tales como DOTMA [Felger et al *Proc. Natl. Acad. Sci USA* (4:7413 (1987))] pueden aumentar la eficacia y facilidad de llevar a cabo este enfoque.

65 6.4 Modelos en animales para CF

La creación de un modelo con ratón u otro animal para CF será crucial para comprender la enfermedad y para ensayar posibles terapias (para una revisión general de la creación de los modelos con animales, véase Erickson, *Am. J. Hum. Genet* 43:582 (1988)]. Actualmente, no existe ningún modelo de animal de la CF. La conservación evolutiva

del gen de CF (según se demuestra mediante los borrones de hibridación de especies cruzadas para E4.3 y H1.6), como se muestra en la Figura 4, indica que existe un gen ortólogo en el ratón (designado de aquí en adelante mCF, y su correspondiente proteína mCFTR) y será posible clonarlo en bancos genómicos y de ADNc de ratón utilizando las sondas del gen de CF humano. Se espera que la generación de una mutación específica en el gen de ratón análoga a la mutación F508 será la más óptima para reproducir el fenotipo, a pesar de que una inactivación completa del gen de mCFTR también será un mutante útil de generar.

A. Mutagénesis. La inactivación del gen de mCF se puede conseguir por mutagénesis química [por ejemplo Johnson et al Proc. Natl. Acad. Sci. USA 78: 3138 (1981)] o de rayos X [Popp et al J. Mol. Biol. 127:141 (1979)] de gametos de ratón, seguido de fertilización. El brote heterozigótico de la inactivación de mCFTR se puede luego identificar mediante transferencia Southern para demostrar la pérdida de un alelo mediante dosificación, o incapacidad de heredar un alelo parental si se está verificando un marcador de RFLP. Este enfoque ha sido utilizado previamente con éxito para identificar mutantes de ratón para una α -globina [Whitney et al Proc. Natl. Acad. Sci. USA 77: 1087 (1980)], fenilalanina hidroxilasa [McDonald et al Pediatr. Res 23:63 (1988)] y anhidrasa carbónica II [Lewis et al Proc. Natl. Acad. Sci. USA 85:1962, (1988)].

B. Transgénicos. Una versión normal o mutante de CFTR o mCFTR se puede insertar en la línea germinal de ratón utilizando ahora técnicas convencionales de la inyección de oocitos [Camper, Trends in Genetics (1988)]; alternatively, si es deseable inactivar o reemplazar el gen de mCF endógeno, se puede aplicar el sistema de recombinación homólogo utilizando células germinales embrionarias (ES- embryonic stem) [Capecchi, Science, 244: 1288 (1989)].

1. Inyección de oocitos. La colocación de una o más copias del gen de mCF normal o mutante en un lugar aleatorio en la línea germinal del ratón se puede conseguir mediante microinyección del pronúcleo de un oocito de ratón recién fertilizado, seguido de la reimplantación en una madre adoptiva pseudo-preñada. A continuación, los ratones nacidos se pueden rastrear en cuanto a integrantes utilizando el análisis del ADN de la cola en cuanto a la presencia de secuencias de genes de CF humanos. Este mismo protocolo se puede utilizar para insertar un gen de mCF mutante. Para generar un modelo de ratón, se desearía colocar este transgen en un fondo de ratón en el que el gen de mCF endógeno ha sido inactivado, ya sea por mutagénesis (véase antes) o por recombinación homóloga (véase a continuación). El transgen puede ser: a) una secuencia genómica completa, aunque el tamaño de ésta (aproximadamente 250 kb) requeriría que fuese inyectado en forma de un cromosoma o un fragmento de cromosoma artificial de levadura; b) un ADNc con el promotor natural o un promotor heterólogo; c) un "minigen" que contenga la totalidad de la región codificante y diversos otros elementos tales como intrones, promotor y elementos flanqueantes 3' que se sabe son necesarios para una expresión óptima.

2. Infección retroviral de embriones tempranos. Esta alternativa implica insertar el gen de CFTR o mCF en un vector retroviral e infectar directamente embriones de ratón en estadios de desarrollo tempranos generando una quimera [Soriano et al Cell 46:19 (1986)]. Al menos algunos de estos conducirán a una transmisión de la línea germinal.

3. Células ES y recombinación homóloga. El enfoque de la célula germinal embrionaria (Capecchi, *supra* y Capecchi, Trends Genet 5:70 (1989)) permite la posibilidad de efectuar una transferencia de genes y luego rastrear las células omnipotentes resultantes para identificar los raros sucesos de recombinación homóloga. Una vez identificados, estos sucesos se pueden utilizar para generar quimeras mediante la inyección de blastocitos de ratón, y una proporción de los ratones resultantes mostrará una transmisión de la línea germinal desde la línea recombinante. Existen varias maneras de que esto sea útil en la generación de un modelo de ratón para CF:

a) La inactivación del gen de mCF se puede conseguir convenientemente diseñando un fragmento de ADN que contenga secuencias procedentes de un exón de mCFTR que flanquea un marcador seleccionable tal como neo. La recombinación homóloga conducirá a la inserción de las secuencias neo en el centro de un exón, inactivando mCFTR. Los sucesos de recombinación homóloga (por normal general, aproximadamente 1 de 1000) se pueden reconocer por los sucesos heterólogos mediante análisis de ADN de clones individuales [habitualmente utilizando PCR, Kim et al Nucleic Acids Res. 16:8887 (1988), Joyner et al Nature 338:153 (1989); Zimmer et al *supra*, pág. 150] o utilizando una selección negativa contra los sucesos heterólogos [tal como el uso de un gen TK de HSV en el extremo de la construcción, seguido de la selección de ganciclovir, Mansour et al, Nature, 336: 348 (1988)]. Este ratón mCFTR inactivado se puede utilizar luego para introducir un gen de CF mutante o un gen de mCF que contiene la anomalía F508 o cualquier otra mutación deseada.

b) Es posible que se puedan crear en una etapa mutantes específicos de mCFTR. Por ejemplo, se puede preparar una construcción que contenga 9 secuencias del intrón de mCF en el extremo 5', un gen neo seleccionable en el centro e intro 9 + exón 10 (que contiene la versión de ratón de la mutación F508) en el extremo 3'. Un suceso de recombinación homóloga conduciría a la inserción del gen neo en el intrón 9 y la sustitución del exón 10 por la versión mutante.

c) Si la presencia del marcador neo seleccionable en el intrón alteró la expresión del gen de mCF, sería posible escindirlo en una segunda etapa de recombinación homóloga.

d) También es posible crear mutaciones en la línea germinal de ratón inyectando oligonucleótidos que contienen la mutación de interés y rastreando las células resultantes por PCR.

ES 2 210 224 T3

Esta realización de la invención ha considerado principalmente un modelo de ratón para la fibrosis quística. La figura 4 muestra una hibridación entre especies no sólo a ADN de ratón, sino también a ADN bovino, de hámster y de pollo. Así, se contempla que también existirán en muchas otras especies un gen ortólogo. Así, se contempla que será posible generar otros modelos con animales utilizando una tecnología similar.

5

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Una molécula de ADN purificada que comprende un gen de fibrosis quística, comprendiendo dicho gen una secuencia de ADN seleccionada del grupo que consiste en:

- (a) secuencias de ADN que corresponden a la secuencia de ADN de la figura 1 desde la posición 1 a la posición 1480 del residuo aminoácido o una forma polimórfica de dicha secuencia de ADN en la que la sustitución de nucleótidos de dicha secuencia de ADN no afecta a la función esencial del polipéptido codificado con ello;
- (b) secuencias de ADN que codifican un polipéptido regulador de la conductancia de la transmembrana de fibrosis quística (CFTR) normal que tiene la secuencia de acuerdo con la figura 1 desde la posición 1 a la posición 1480 de un residuo aminoácido o una forma polimórfica de dicho polipéptido, en el que la sustitución de aminoácidos en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta al funcionamiento esencial de la misma o su perfil hidropático o estructura secundaria o terciaria;
- (c) secuencias de ADN que corresponden a un fragmento de la secuencia de ADN de la figura 1 que incluyen al menos 16 nucleótidos secuenciales entre las posiciones 134 y 4573 de la secuencia de nucleótidos y que pueden rastrear y detectar la presencia de un gen o producto génico de la fibrosis quística normal o defectuoso;
- (d) secuencias de ADN que comprenden al menos 16 nucleótidos y que codifican un fragmento de la secuencia de aminoácidos de la figura 1 que pueden rastrear y detectar la presencia de un gen o producto génico de la fibrosis quística normal o defectuoso; y
- (e) secuencias de ADN que codifican un epítipo codificado por al menos 18 nucleótidos secuenciales en la secuencia de ADN de la figura 1 entre las posiciones 1 y 1480 del residuo aminoácido.

2. Una molécula de ADN purificada de (a), (b) o (c) según la reivindicación 1, que contiene, además, al menos una mutación en la secuencia de ADN que, si se expresa en células del cuerpo humano, se asocia con la función celular alterada que se correlaciona con la fibrosis quística, en donde dicha mutación incluye una delección de tres nucleótidos que codifican fenilalanina en la posición 508 del residuo aminoácido.

3. Una sonda de ácido nucleico purificado que comprende una secuencia de nucleótidos de ADN o ARN que corresponde a la secuencia de acuerdo con las características (c), (d) o (e) según la reivindicación 1 o 2.

4. Una sonda de ácido nucleico según la reivindicación 3, en donde dicha secuencia comprende AAA GAA AAT ATC ATC TTT GGT GTT, y su complemento.

5. Un vector de clonación recombinante que comprende una molécula de ADN de acuerdo con la reivindicación 1 o 2.

6. Un huésped no humano, transformado con el vector de acuerdo con la reivindicación 5.

7. Un método para producir un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) normal o mutante, que comprende las etapas de:

- (a) cultivar una célula huésped transfectada con el vector de la reivindicación 5 en un medio y en condiciones favorables para la expresión de un polipéptido CFTR normal o mutante; y
- (b) aislar el polipéptido CFTR normal o mutante expresado.

8. Una molécula de ADN purificada que comprende una secuencia de ADN que codifica un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) mutante que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del residuo aminoácido, o una forma polimórfica de dicho polipéptido en la que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria, **caracterizado** además por una delección de tres nucleótidos que codifican fenilalanina en la posición 508 del residuo aminoácido.

9. Un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) purificado, **caracterizado** por tener un peso molecular del péptido de aproximadamente 170.000 dalton, seleccionado del grupo que consiste en:

- (a) un polipéptido CFTR normal que tiene actividad que afecta a la conductancia de iones de la transmembrana de la célula y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido o una forma polimórfica de dicho polipéptido CFTR normal, en que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria; y

ES 2 210 224 T3

(b) un polipéptido CFTR mutante que tiene actividad de fibrosis quística en células humanas y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido, excepto por la delección de fenilalanina en la posición 508 del residuo aminoácido.

5 10. Un polipéptido codificado por la expresión de una secuencia de ADN de acuerdo con la reivindicación 1 o 2, exhibiendo dicho polipéptido la actividad inmunológica de polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) normal o mutante.

10 11. Un procedimiento para aislar la proteína CFTR normal o mutante de acuerdo con la reivindicación 10 a partir de células que contienen dicha proteína, que comprende las etapas de:

(a) solubilizar la proteína de una membrana celular seleccionada en la que se expresa dicha proteína CFTR normal o mutante, para proporcionar una solución de dicha proteína CFTR;

15 (b) separar dicha proteína CFTR de dicha solución poniendo en contacto dicha solución con anticuerpos contra dicha proteína CFTR normal o mutante, estando dichos anticuerpos inmovilizados sobre un sustrato,

(c) aclarar dicho sustrato para separar proteína no adherida a dichos anticuerpos;

20 (d) liberar dicha proteína CFTR de dichos anticuerpos para aislar con ello dicha proteína CFTR, y

(e) purificar dicha proteína CFTR para separar cualquier otra proteína de mamífero remanente.

25 12. Un método de rastrear a un sujeto para determinar si dicho sujeto es un portador de fibrosis quística (CF) o un paciente de CF, método que comprende:

proporcionar un ensayo para detectar en una muestra biológica a rastrear la presencia de al menos un miembro procedente del grupo que consiste en un gen de CF normal, productos génicos de CF normal, un gen de CF mutante, productos génicos de CF mutantes y sus mezclas, un polipéptido CFTR normal con una actividad que afecta a la conductancia de iones en la transmembrana de la célula y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 480 del aminoácido o una forma polimórfica de dicho polipéptido CFTR normal en que la sustitución de aminoácidos en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria, un polipéptido CFTR mutante con actividad de fibrosis quística en células humanas y que tiene la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 de aminoácidos excepto la delección de fenilalanina en la posición 508 del residuo aminoácido, en donde el ensayo es

(i) un ensayo que implica hibridación, o

40 (ii) un ensayo inmunológico,

y/o se consigue utilizando una o más hibridaciones que utilizan oligonucleótidos específicos, secuenciación de ADN directa, enzimas de restricción, discriminación sobre la base de la movilidad electroforética en geles con agente desnaturante, protección de RNasa, escisión química o el proceso de detección mediado por ligasa.

45 13. Un método de acuerdo con la reivindicación 12, en el que el ensayo es un ensayo que implica hibridación y la muestra biológica incluye al menos parte del genoma del sujeto.

50 14. Un método de acuerdo con la reivindicación 13, en el que el ensayo comprende, además, una sonda de nucleótido marcada de acuerdo con la reivindicación 3.

15. Un método de acuerdo con la reivindicación 14, en el que dicha sonda comprende la secuencia de nucleótidos de la reivindicación 4.

55 16. Un método de acuerdo con la reivindicación 12, en el que el ensayo es un ensayo inmunológico y la muestra biológica incluye un polipéptido regulador de la conductancia de la transmembrana de la fibrosis quística (CFTR) normal del sujeto y/o un polipéptido CFTR mutante del sujeto: teniendo dicho polipéptido CFTR normal la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido o una forma polimórfica de dicho polipéptido en la que la sustitución del aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria; y teniendo dicho polipéptido CFTR mutante la secuencia de acuerdo con la figura 1 para las posiciones 1 a 1480 del aminoácido, excepto la delección de fenilalanina en la posición 508 del residuo aminoácido.

65 17. Un método de acuerdo con la reivindicación 16, en el que el ensayo incluye, además, un anticuerpo específico para un polipéptido CFTR normal.

18. Un método de acuerdo con la reivindicación 16, en el que el ensayo incluye, además, un anticuerpo específico para un polipéptido CFTR mutante.

ES 2 210 224 T3

19. Un método de acuerdo con la reivindicación 17 o 18, en el que el anticuerpo es al menos un anticuerpo monoclonal.

20. Un método de acuerdo con la reivindicación 12, en el que sujeto es un feto humano en el útero.

21. Un método de acuerdo con la reivindicación 14 o 15, en el que el análisis incluye, además, al menos una sonda de nucleótido adicional de acuerdo con la reivindicación 3, con la condición de que cada sonda de nucleótido en dicho análisis sea diferente.

22. Un procedimiento para rastrear un portador o paciente de fibrosis quística (CF) potencial para identificar la presencia de una mutación de la fibrosis quística identificada en el gen de CF que codifica un polipéptido regulador de la conductancia en la transmembrana de la fibrosis quística (CFTR) que tiene la secuencia de acuerdo con la figura 1 desde la posición 1 a la posición 1480 del aminoácido o una forma polimórfica de dicho polipéptido, en cuya sustitución de aminoácido en las regiones variables de dicho polipéptido con la secuencia de acuerdo con la figura 1 no afecta a su funcionamiento esencial, o su perfil hidropático o estructura secundaria o terciaria, incluyendo dicho procedimiento las etapas de:

(a) hibridar una sonda de ADN de acuerdo con la reivindicación 3 sobre ADN genómico aislado procedente de dicho portador de CF potencial o dicho paciente potencial, abarcando dicha sonda de ADN dicha mutación de fibrosis quística en dicho gen de CF, en donde dicha sonda de ADN es capaz de detectar dicha mutación de la fibrosis quística; y

(b) tratar dicho ADN genómico para determinar la presencia de dicha sonda de ADN y, con ello, indicar, de acuerdo con una manera predeterminada de hibridación, la presencia o ausencia de dicha mutación de fibrosis quística.

23. Un procedimiento para detectar portadores o pacientes de fibrosis quística, en donde dicho procedimiento consiste en poner en contacto un gen de CF del paciente o portador con una endonucleasa de restricción, determinar la presencia o ausencia de un sitio de endonucleasa de restricción en el gen de CF y comparar el modelo de los sitios de endonucleasa de restricción en el gen de CF con un gen de CF normal que corresponde a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1, la presencia de sitios adicionales al gen de CF normal o la ausencia de sitios en el gen de CF normal indicativos de un gen de CF mutante.

24. Un procedimiento para detectar portadores de fibrosis quística, en donde dicho procedimiento consiste en determinar una movilidad diferencial de productos de PCR heteroduplex en geles de poliacrilamida como resultado de inserciones o deleciones en el gen de la fibrosis quística (CF) mutante comparado con el gen de CF normal correspondiente a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1.

25. Un estuche para someter a ensayo, mediante inmunoensayo, la presencia de un gen de fibrosis quística (CF) normal correspondiente a la secuencia de nucleótidos que codifica al polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1 o un gen de CF mutante que tiene al menos una deleción de tres pares de bases en el gen de CF normal, deleción que da como resultado la deleción de una fenilalanina procedente de un residuo aminoácido de la posición 508 de la figura 1, que comprende:

(a) un anticuerpo que se une específicamente a un producto génico del gen de CF normal o el gen de CF mutante de acuerdo con la reivindicación 9;

(b) medios de reactivos para detectar la unión del anticuerpo al producto génico; y

(c) estando los medios de anticuerpo y reactivos cada uno presentes en cantidades eficaces para llevar a cabo el inmunoensayo.

26. Un estuche para someter a ensayo la presencia de un gen de fibrosis quística (CF) normal correspondiente a la secuencia de nucleótidos que codifica el polipéptido entre el residuo aminoácido en la posición 1 y 1480 de la figura 1 o un gen de CF mutante que tiene al menos una deleción de tres pares de bases en el gen de CF normal, deleción que da como resultado la deleción de una fenilalanina procedente del residuo aminoácido en la posición 508 de la figura 1, mediante hibridación, que comprende:

(a) una sonda de oligonucleótido de acuerdo con la reivindicación 3 que se une específicamente al gen de CF normal o al gen de CF mutante;

(b) medios de reactivos para detectar la hibridación de la sonda de oligonucleótido al gen de CF normal o al gen de CF mutante; y

(c) estando presentes cada uno de los medios de sonda y reactivos en cantidades eficaces para llevar a cabo el ensayo de hibridación.

ES 2 210 224 T3

27. Un anticuerpo regulador de la conductancia en la transmembrana anti-fibrosis quística (CFTR), policlonal o monoclonal, inmunológicamente activo, específico para un polipéptido CFTR de acuerdo con la reivindicación 9.

5 28. Un hibridoma que produce un anticuerpo monoclonal específico para un polipéptido regulador de la conductancia en la transmembrana de la fibrosis quística (CFTR) de acuerdo con la reivindicación 9.

10 29. Una composición para el tratamiento de la fibrosis quística en un paciente que comprende una cantidad terapéuticamente eficaz de una proteína de acuerdo con los rasgos (a) de la reivindicación 9, en combinación con un vehículo farmacéuticamente aceptable.

30. Una composición de acuerdo con la reivindicación 29, en donde dicho vehículo incluye una proteína tensioactiva de los pulmones para facilitar la aplicación de dicha composición a células epiteliales respiratorias.

15 31. Una composición para la terapia génica de la fibrosis quística, que comprende una molécula de ADN de acuerdo con la reivindicación 1 y un vehículo para suministrar dicha molécula de ADN a una célula de un paciente de fibrosis quística.

32. Una composición de acuerdo con la reivindicación 31, en donde dicho vehículo es un vector recombinante.

20 33. Un animal no humano que comprende un sistema de células heterólogo que comprende un vector de clonación recombinante de acuerdo con la reivindicación 5, vector que induce síntomas de fibrosis quística en dicho animal.

25 34. Un ratón transgénico que exhibe síntomas de fibrosis quística, que se puede obtener mediante las etapas que comprenden introducir un gen de fibrosis quística de acuerdo con la reivindicación 2 que codifica una proteína reguladora de la conductancia en la transmembrana de la fibrosis quística mutante que incluye al menos una delección de fenilalanina en la posición correspondiente a 508 en la figura 1 en un oocito de ratón, o un embrión de ratón; y, opcionalmente, inactivar el gen de la fibrosis quística de ratón endógeno.

30

35

40

45

50

55

60

NOTA INFORMATIVA: Conforme a la reserva del art. 167.2 del Convenio de Patentes Europeas (CPE) y a la Disposición Transitoria del RD 2424/1986, de 10 de octubre, relativo a la aplicación del Convenio de Patente Europea, las patentes europeas que designen a España y solicitadas antes del 7-10-1992, no producirán ningún efecto en España en la medida en que confieran protección a productos químicos y farmacéuticos como tales.

65

Esta información no prejuzga que la patente esté o no incluida en la mencionada reserva.

FIG.1.

1 AATTGGAAGCAAAATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCA

61 GAGTAGTAGGCTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGC

121 GCCCGAGAGACCATGCAGAGGTGGCCTCTGGAAAAGGCCAGCGTTGTCTCCAAACTTTTT
M O R S P L E K A S V V S K L F 16

181 TTCAGCTGGACCAGACCAATTTGAGGAAAGGATACAGACAGCGCCTGGAAATGTCAGAC
F S W T R P I L R K G Y R Q R L E L S D 36

241 ATATACCAAATCOCTTCTGTTGATTCTGCTGACAACTATCTGAAAAATGGAAAGAGAA
I Y Q I P S V D S A D N L S E K L E N E 56

301 TGGGATAGAGAGCTGGCTTCAANGAAAATCCTAAACTCATTAAATGCCCTTCGGCGATGT
W D R E L A S K K N P K L I N A L R R C 76

361 TTTTCTGGAGATTATGTTCTATGGAATCITTTTATATTTAGGGGAAGTCACCAAAGCA
F F W R F H F Y G I F L Y L G E V T K A 96

421 GTACAGCCTCTCTTACTGGGAAGAATCATAGCTTCCATGACCCGGATAACAAGGAGGAA
V Q P L L L L G R I I A S Y D P D N K E E 116

481 CGCTCTATCCGGATTATCTAGGCATAGCCTTATGCCCTTCTCTTATTGTGAGGACACTG
R S I A I Y L G I G L C L L F I V R T L 136

541 CTCTACACCCAGCCATTTTTGGCCTTCATCACATGGGAATGCAGATGAGAAATAGCTATG
L L H P A I P G L N N I G H Q H R I A H 156

601 TTTAGTTGATTATAAGAAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAATAAGT
F S L I Y K K T L K L S S R V L D K I S 176

661 ATTGGACAACITGTTAGTCTCCTTTCCAAACCTGAACAAAATTTGATGAAGGACTTGCA
I G O L V B L L S N N L N X P D E G L A 196

721 TTGGCACATTTCTGTGGATCCCTCTTGCAGTGGCACTCTCATGGGCTAATCTGG
L A R F V M I A P L O V A L L M G L I N 216

781 GAGTTGTTACAGCGGCTCGCCTTCTGTGGACTTGGTTTCTGTATAGTCTTGCCTTTT
E L L O A S A P F C G L G F L I V L L L F 236

841 CAGGCTGGGCTAGGGAGAATGATGATGAAGTACAGAGATCAGAGAGCTGGGAAGATCAGT
O G L G R N H M K Y R D O R A G K I S 256

901 GAAAGACTTGTGATTACCTCAGAAATGATTGAAAATATCCAAATCTGTTAAGGCATACTGC
E R L V I T S E N I E N I O S V K A Y C 276

961 TGGGAAGAAGCAATGGAAAAATGATTCAAACCTTAAGACAAACAGAACTGAAACTGACT
W E E A H E K M I E N L R Q T E L K L T 296

1021 CGGAAGGCAGCCTATGTGAGATACTTCAATAGCTCAGCCTTCTTCTCAGGGTCTTT
R K A A Y V R Y F N S S A F P F S G F F 316

1081 GTGGTGTTTTATCTGTGCTTCCCTATGCACATAATCAAAGGAATCATCCTCCGGAAAATA
V V P L S V L P Y A L I N G I I L R K I 336

1141 TTCACCACCATCTCATTTCTGCATTGTTCTGGCATGGCGGTCACTCGGCAATTTCCCTGG
F T T I S F C I V L R M A V T R Q F P W 356

1201 GCTGTACAACATGGGTATGACTCTCTTGGAGCAATAACAAAATACAGATTTCTTACAA
A V O T W Y D S L G A I N K I O D P L O 376

1261 AAGCAAGAATAAAGACATGGAATATAACTTAACGACTACAGAAGTACTGATGGAGAAT
K Q E Y K T L E Y N L T T T E V V M E N 396

1321 GTAACAGCCTTCTGGGAGGAGGATTTGGGGAATATTTGAGAAAGCAAACAAAACAAT
V T A F W E E G F G E L F E K A K O N N 416

FIG.1 (cont.)

1381 N M R K T S N S D D S L F F S N P S L L 436
 AACAAATAGAAAACCTCTAATGGTGATGACAGCCTCTCTTCAGTAAATTTCTCACTTCTT
 1441 G T P V L K D I N F K I E R G Q L L A V 456
 GGTACTCCTGTCCTGAAAGATATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTT
 1501 A G S T Q A Q K I T E L L L M N I N G E L E 476
 GCTGGATCCACTGGAGCAGCCAAACTTCACTTCTAATGATGATTATGGGAGAAGCTGGAG
 1561 P S E Q K I K H S G R I S F C S O P E W 496
 CCTTCAGAGGGTAAAATTAAGCACAGTGGAAAGAAATTCATTCTGTTCTCAGTTTTCCTGG
 1621 I M P G T I K E N I I F G V S Y D E T R 516
 ATTATCCCTGGCACCATTAAAGAAATATCATCTTTGGTGTTCCTATGATGAATATAGA
 1681 Y R E V I K A C Q L E E I D I S K F A E K 536
 TACAGAAGCGTCATCAAAGCATGCCAACTAGAAAGAGACATCTCCAAGTTTGCAGAGAAA
 1741 D N I V L G E Q G I T L S G G O R A R I 556
 GACAATATAGTTCCTGGAGAAGGTGGAAATCACACTGAGTGGAGGTCACGAGCAAGAATT
 1801 S L A N A V Y K D A D L Y L L D S P F G 576
 TCTTAGCAAGAGCAGTATACAAAGATGCTGATTGCTATTTATTAGACTCTCCTTTTGG
 1861 Y L D V L E E E E I F E H C V C K L M A 596
 TACCTAGATGTTTAAACAGAAAAGAAATATTTCAAGCTGTGTCTGTAAACTGATGGCT
 1921 N K T R I L V T S K M E N L K K A D K I 616
 AACAAAACCTAGGATTTTGGTCACTTCTAAAATGGAAACATTTAAAGAAAGCTGACAAAATA
 1981 L I L N E G S S Y F Y G T F S E L O W L 636
 TTAATTTTGAATGAAGGTAGCAGCTATTTTATGGGACATTTTCAGAACTCCAAAATCTA
 2041 Q P D F S S K L M G C D S F D O P S A Z 656
 CAGCCAGACTTTAGCTCAAAACTCATGGGATGTGATTCTTTCGACCAATTTAGTGCAGAA
 2101 R R N S I L T E T L H R F S L E G D A P 676
 AGAAGAANTTCAATCCTAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCT
 2161 V S W T E T K K O S F K O T G E F G E K 696
 GTCTCCTGGACAGAAACAAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGCGAAAAA
 2221 R K N S I L N P I N S I R K F S I V O K 716
 AGGAAGAATTCATTTCTCAATCCAATCAACTCTATACGAAAATTTTCCATTGTGCAAAAG
 2281 T P L O M N G I E E D S D E P L E R R L 736
 ACTCCCTTACAAATGAATGGCATCGAAGAGGATTCTGATGAGCCTTTACAGAGAAGGCTG
 2341 S L V P D S E O G E A I L P R I S V I S 756
 TCCTTAGTACCAGATTCTGAGCAGGGAGAGGCGATACTGCTCCCATCAGCGTGCATCAGC
 2401 T G P T L O A R R R O S V L N L N T H S 776
 ACTGGCCCCACGCTTCAGGCACGAAGGAGGCACTCTGCTCCTGAACCTGATGACACACTCA
 2461 V N Q G Q N I H R K T T A S T R K V S L 796
 GTTAACCAAGGTGAGAACATTCACCGAAAAGCAACAGCATCCACACGAAAAGTCTCACTG
 2521 A P Q A N L T E L D I Y S R R L S O E T 816
 GCCCTCAGGGCAACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAACT

FIG.1 (cont.)

2581 G L E I S E E I N E E D L N E C L F D D 836
 GGCTTGC^{AA}ATAAGTGAAGAAATTAACGAAGAAGACTTAAACGAGTGCCTTTTTGATGAT

2641 M E S I P A V T T W N T Y L R Y I T V H 856
 ATGGAGAGCATAACCAGCAGTGACTACATGGAACACATACCTTCGATATATTACTGTCCAC

2701 K S L I P V L I M C L V I F L A E V A A 876
 AAGAGCTTAATTTTGTGCTAATTTGGTGTCTAGTAATTTTCTGCCAGAGGTGGCTGCT

2761 S L V V L M L L G N T P L O D K G N S T 896
 TCTTTGGTGTGCTGTGGCTCCTTGGAAACACTCCTCTTCAAGACAAGGGAAATAGTACT

2821 H S R N N S Y A V I I T S T S S Y Y V F 916
 CATAGTAGAAATAACAGCTATGCAGTGATTATCACAGCACCAGTTCGTATTATGTGTTT

2881 V I Y V G V A D T L L A M G F F R G L P 936
 TACATTTACGTGGAGTAGCCGACACTTGTCTGCTATGGGATTCTTCAGAGGTCTACCA

2941 L V H T L I T V S K I L H H K M L H S V 956
 CTGGTGCATACTCTAATCACAGTGTGGA^{AA}ATTTTACACCACAA^{AA}ATGTTACATCTGT

3001 L Q A P N S T L N T L K A F G I L N R F 976
 CTTCAAGCACCTATGTCAACCTCAACACGTTGAAAGCAGGTGGGATTCTTAATAGATTC

3061 S K D I A I L D D L L P L T I F D F I O 996
 TCCAAAGATATAGCAATTTGGATGACCTTCTGCTCTTACCATATTTGACTTCATCCAG

3121 L L L I V I G A I A V V A V L O P Y I F 1016
 TTGTTATTAATTTGTGATTGGAGCTATAGCAGTTGTCGAGTTTTACAACCTACATCTTT

3181 V A T V P V I V A F I M L R A Y F L O T 1036
 GTTGCACAGTGCAGTGTATAGTGGCTTTTATTATGTTGAGAGCATATTTCTCCAAACC

3241 S O O L K O L E E E G R E P I F T H L V 1056
 TCACAGCAACTCAACA^{AA}CTGGAATCTGAAGGCAGGAGTCCAATTTCACTCATCTTGT

3301 T E L K G L W T L R A F G R O P Y F E T 1076
 ACAAGCTTAA^{AA}AGGACTATGGACACTTCTGCTCTGGACGGCAGCCTTACTTTGAAACT

3361 L F H K A L N L E T A N W F L Y L S T L 1096
 CTGTTCCACAAGCTCTGAATTTACATACTGCCA^{AA}CTGGTTCTTGTACCTGTCAACACTG

3421 R W F Q M R I E M I F V I F F I A V T R 1116
 CGCTGGTCCA^{AA}ATGACAATAGAAATGATTTTCTCATCTTCTTCA^{AA}TGCTGTACCTTC

3481 I S I L T T G E G E G H V G I I L P L A 1136
 ATTTCCATTTTAACAACAGGAGAAGGAGAAGGAAGAGTTGGTATTATCCTGACTTTAGCC

3541 M N I M S T L O N A V N E S I D V D S L 1156
 ATGAATATCATGAGTACATTGCAGTGGGCTGTA^{AA}CTCCACCATAGATGTGGATAGCTT

3601 M R S V S R V P K F I D M P T E G K P T 1176
 ATGCCATCTGTGAGCCGAGTCTTTAAGTTCATTGACATGCCAACAGAAGGTAAACCTACC

3661 K S T K P Y K M G O L S K V H I E N S 1196
 AAGTCAACCA^{AA}ACCATAACAAGAAATGCCA^{AA}CTCTCGAAAGTTATGATTATGAGAATTCA

3721 H V K K D D I W P S G G Q M T V K D L T 1216
 CACGTGAAGAAAGATGACATCTGGCCCTCAGGGGCCAAATGACTGTCAAAGATCTCACA

3781 A K Y T E G G N A I L E N I S Y S I S P 1236
 GCAAAATACACAGAAGGTGGA^{AA}ATGCCATATTAGAGAACATTCTTCTCAATAAGTCTT

3841 S O R V G L L G R T S G K E T L L A 1256
 GGCCAGAGGTGGCCCTCTTGGGAAGA^{AA}CTGGATCAGGGAAAGAGTACTTTGTTATCAGCT

FIG.1. (cont.)

3901 F L R L L W F E G E I O I D G V S W D E 1276
 TTTTGAGACTACTGAACACTGAAGGAGAAATCCAGATCGATGGTGTCTTGGGATTCA
 3961 I T L O O W R K A F G V I P Q K V F I E 1296
 ATAACCTTGC AACAGTGGAGGAAGCCCTTGGAGTGATACCACAGAAAGTATTTATTTT
 4021 S G T P R K N L D P F E O W S D O E I W 1316
 TCTGGAACATTTAGAAAAAACTTGGATCCCTATGAACAGTGGAGTGATCAAGAAATATGG
 4081 K V A D E V G L R S V I E O F P G K L D 1336
 AAAGTTCAGATGAGTGGGCTCAGATCTGTGATAGAACAGTTTCTGGGAAGCTTGAC
 4141 F V L V D S G C V L S E G R K O L N C L 1356
 TTTGCTCTGTGGATGGGGCTGTGTCTTAAGCCATGGCCACAAGCAGTTGATGTGCTTG
 4201 A R S V L S K A K I L L L D E P S A E L 1376
 GCTAGATCTGTCTCAGTAAGCCGAAGATCTTCTGCTTGTGATGAACCCAGTGTCTCATTG
 4261 D P V T Y O Y I R R T L K O A P A D C T 1396
 GATCCAGTAACATACCAATAATAGAAAGACTCTAAAACAAGCATTGCTGATTGCACA
 4321 V I L C E N R I E A M L E C O O F L V I 1416
 GTAATCTCTGTGAACACAGGATAGAAGCAATGCTGGAATGCCAACAAATTTTGTGTCATA
 4381 E E N K V R O Y D S I O K L L N E R S L 1436
 GAAGAGAACAAGTGGGAGTACGATTCCATCCAGAACTGCTGAACGAGAGGAGCCCTC
 4441 F R Q A I S P S D R V K L P P R R N S S 1456
 TTCGGCAAGCCATCAGCCCTCCGACAGGGTGAAGCTCTTCCCCACCGAAGCTCAAGC
 4501 K C K S K P Q I A A L K E E T E E V O 1476
 AAGTGCAGTCTAAGCCAGATTGCTGCTCTGAAAGAGGAGACAGAAGAAGAGGTGCAA
 D T R L = 1480
 4561 GATACAAGGCTTTAGAGAGCAGCATAAATGTTGACATGGGACATTTGCTCATGGAATTGG
 4621 AGCTCGTGGGACAGTCACTCATGGAATTGGAGCTCGTGGAACAGTTACCTCTGCCTCAG
 4681 AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAGAAACATTTGGTAAGGGCAATTGAGC
 4741 ACACGATATGGGCTCTGATAAATGGCTTCTGGCAATAGTCAAATTTGTGTGAAGGTAC
 4801 TTCAAATCTTCAAGATTTACCACTTGTGTTTTGCAAGCCAGATTTCTCTGAAACCCCTT
 4861 GCCATGTCTAGTAATTTGAAAGGCAGCTCTAAATGTCAATCAGCCTAGTGTATCAGCTT
 4921 ATTCTCTAGTCAAACCTGTTAATTTGTAGTCTTGCACAACTGAATCATACTTCTTA
 4981 GGGTTATGATTAAGTAACTGATAACTGCAAACTTCAGCGGTTTATATAAGCTTGTATTCTT
 5041 TTTTCTCTCTCTCCCATGATGTTTACAACACAACTATATTGTTTGTAAAGCAATTCCA
 5101 ACTATCTCATTTCCAAGCAAGTATTAGAATACCACAGGAACCACAGACTGCACATCAA
 5161 ATATGCCCAATTCACATCTAGTGAAGCAGTCAAGAAAGAGAATCTCCAGATCTGGAAAT
 5221 CAGGTTAGTATTGTCCAGGTCTACCAAAAATCTCAATTTTCAGATAATCACAATACAT
 5281 CCTTACCTGGGAAAGGGCTGTTAATACTTTCAAGGGACAGGATGGTTCCCTTGATG
 5341 AAGAAGTTGATATGCCTTTTCCCACTCCAGAAAGTGACAAGCTCACAGACCTTTGAACT
 5401 AGAGTTTAGCTGGAAAAGTATGTTAGTGCAAATTTGTCACAGGACAGCCCTTCTTTCCACA
 5461 GAAGCTCCAGGTAGAGGGTGTGTAAGTAGATAGGCCATGGGCACTGTGGGTAGACACACA
 5521 TCAGTCCAAGCATTAGATGATAGGTTGATGGTGGTATGTTTTCAGGCTAGATGATG
 5581 TACTTCATGCTGTCTACACTAAGA GAGAATGAGAGACACACTGAAGAAGCACCAATCATG
 5641 AATTAGTTTTATATGCTTCTGTTTTATAATTTGTGAAGCAAAAATTTTCTCTAGGAAA
 5701 TATTTATTTTAAATAATGTTTCAAACATATATACAATGCTGTATTTTAAAGAAATGATTA
 5761 TCAATTACATTTGTATAAAATAATTTTTATATTTGAAATATTGACTTTTTATGGCACTAG
 5821 TATTTTATGAATAATTTATGTTTAAACTGGGACAGGGGAGAACCTAGGGTGATATTAACC
 5881 AGGGCCATGAATCACCTTTTGGTCTGGAGGAAAGCCCTTGGGGCTGATGAGTGTGTC
 5941 CACAGCTGTATGATTTCCAGCCAGACACAGCTTCTAGATGACCTTCTCAAGAAATGCT
 6001 ACCACCAGTCTGACTGTTTCCATCAAGGCTACACTGCTTCTCAACTCCAAACTGACTCT
 6061 TAAGAAGACTGCATTAATTTATTAAGTGAAGAAAATACACTTGTCAATAAAATCCATA
 6121 CATTGTGT (A) n

FIG.2. (cont.)

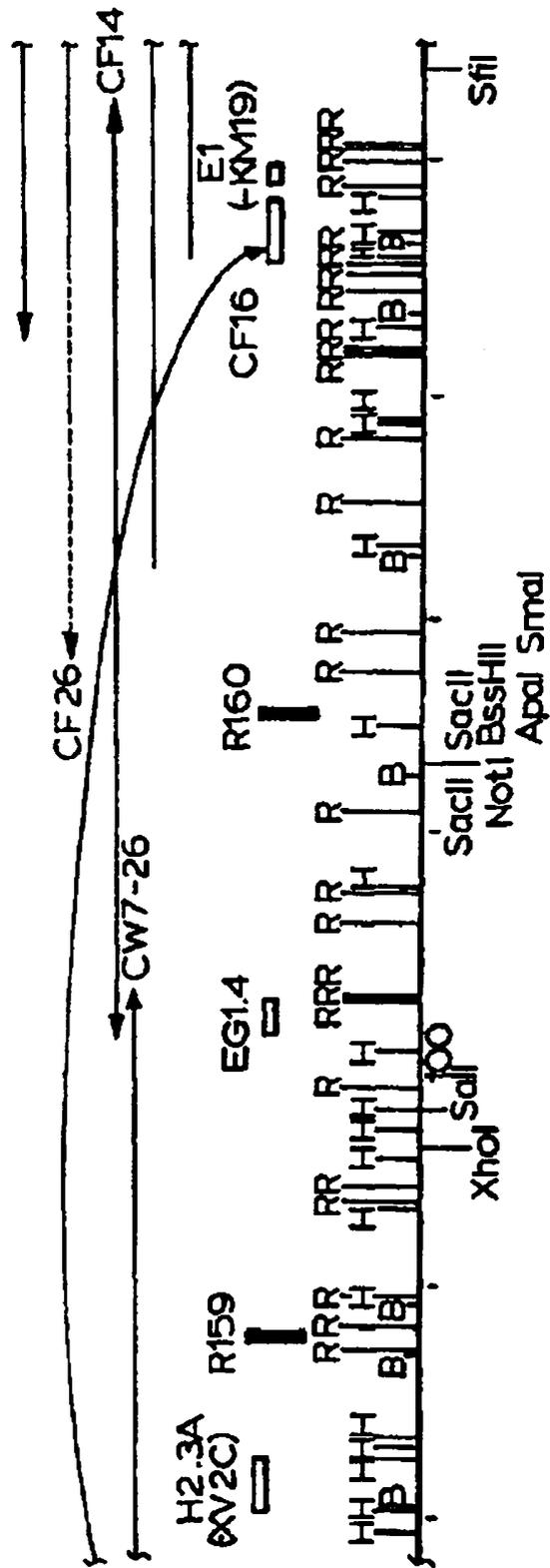


FIG.2 (cont.)

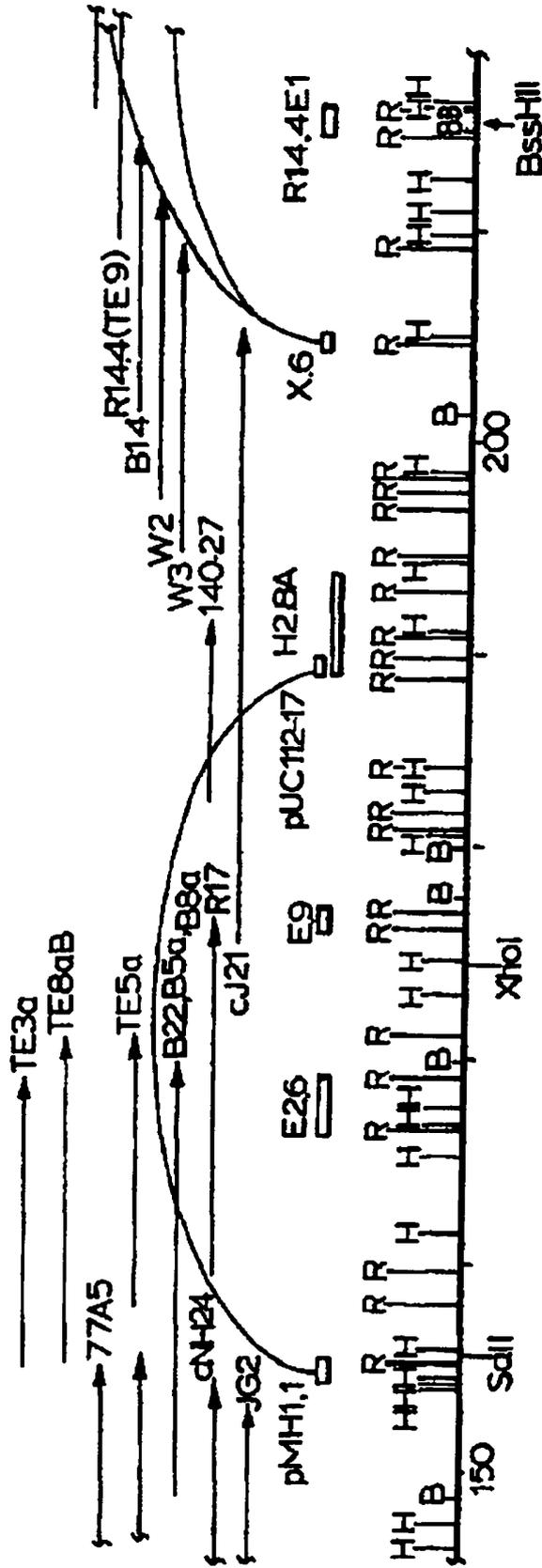


FIG.2 (cont.)

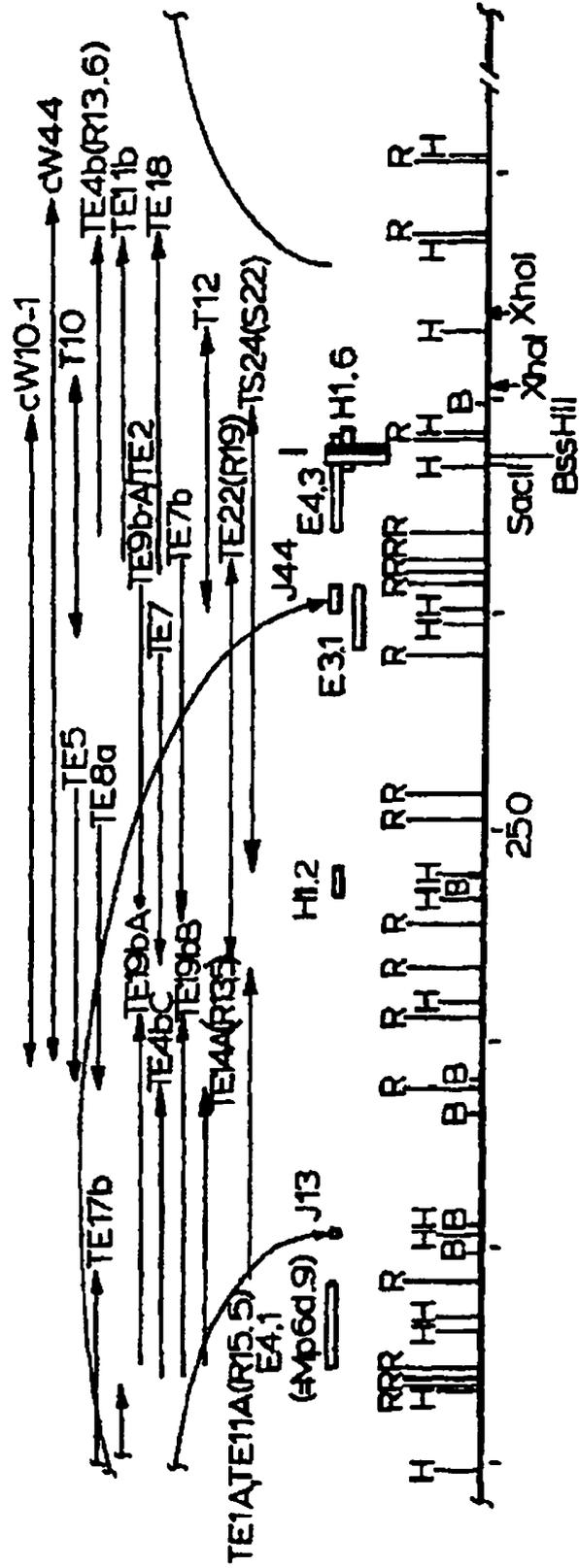


FIG.2 (cont.)

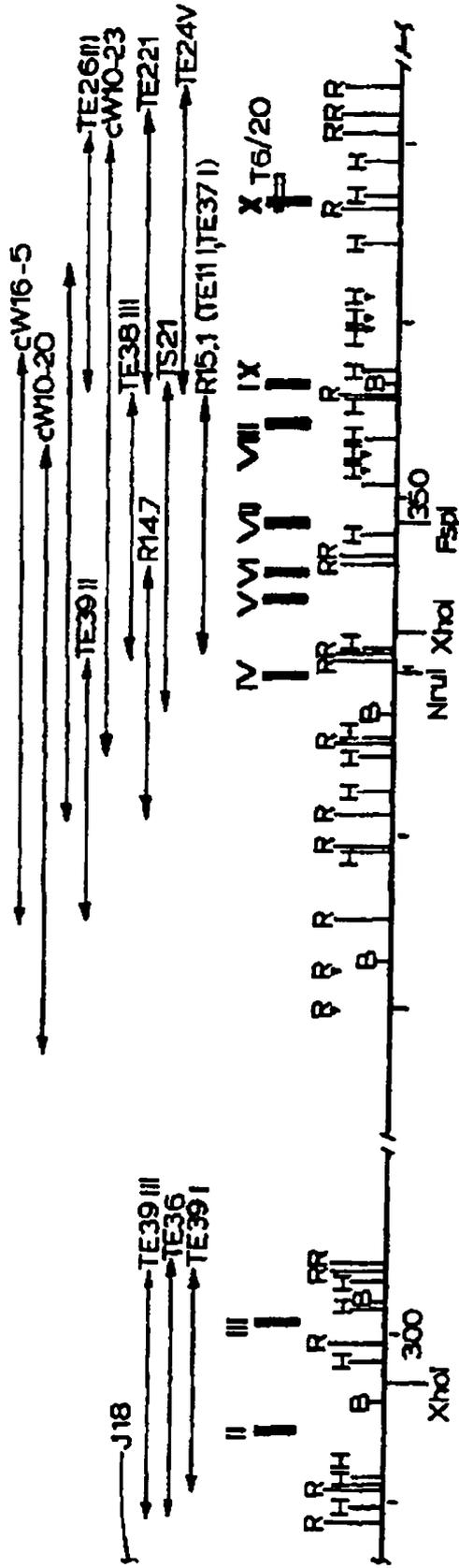


FIG.2 (cont.)

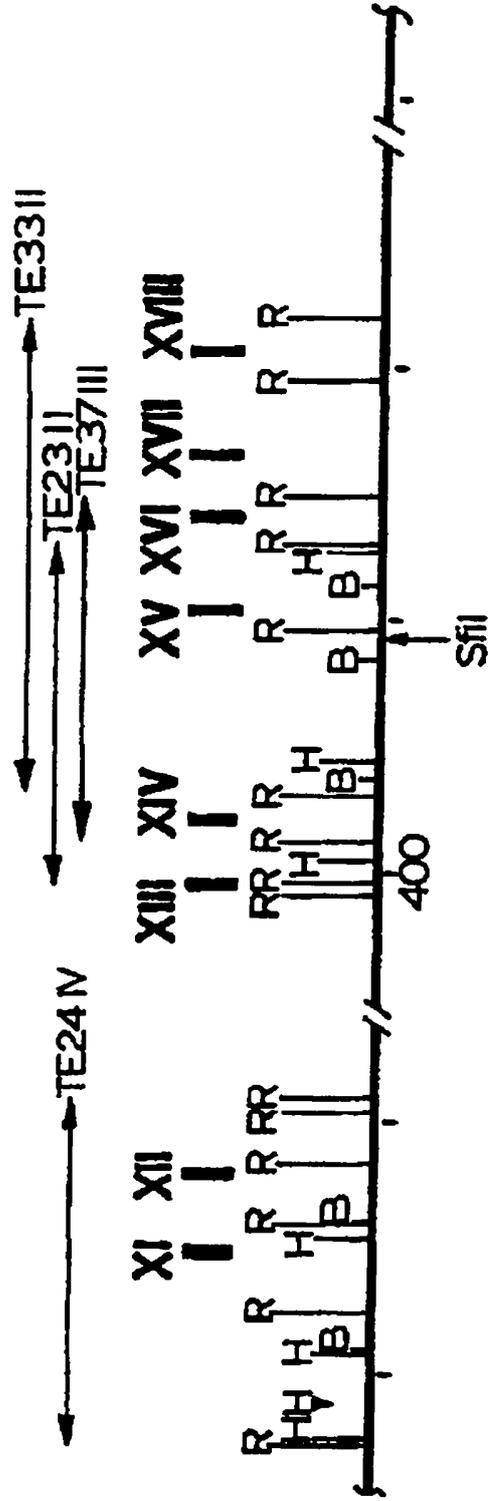
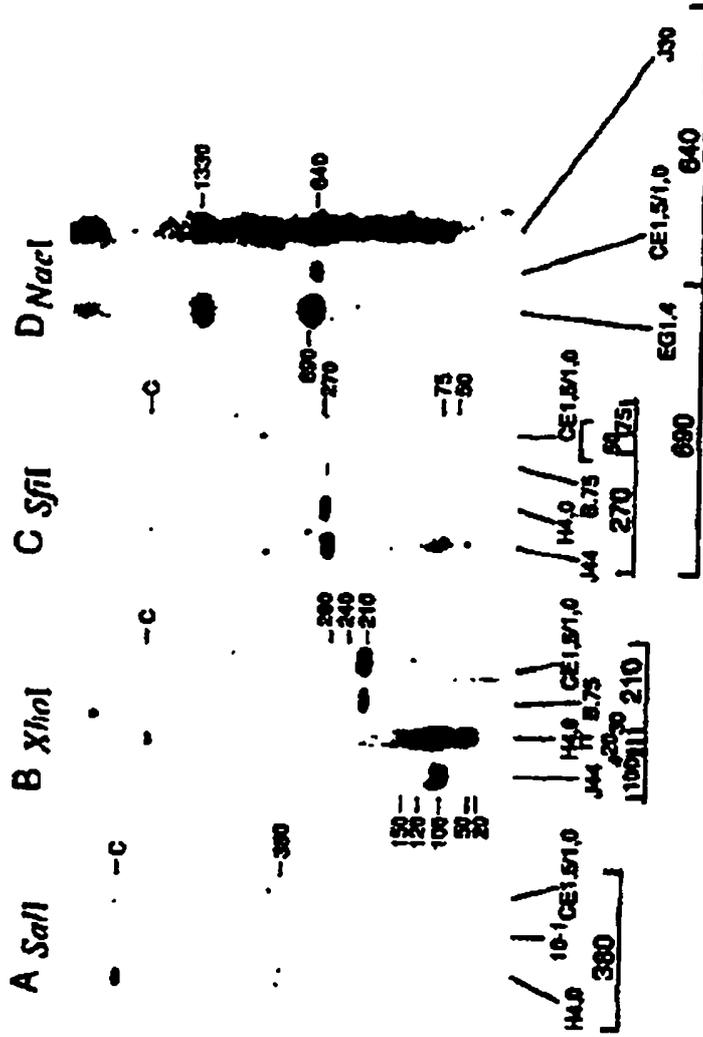


FIG.2 (cont.)



FIG. 3



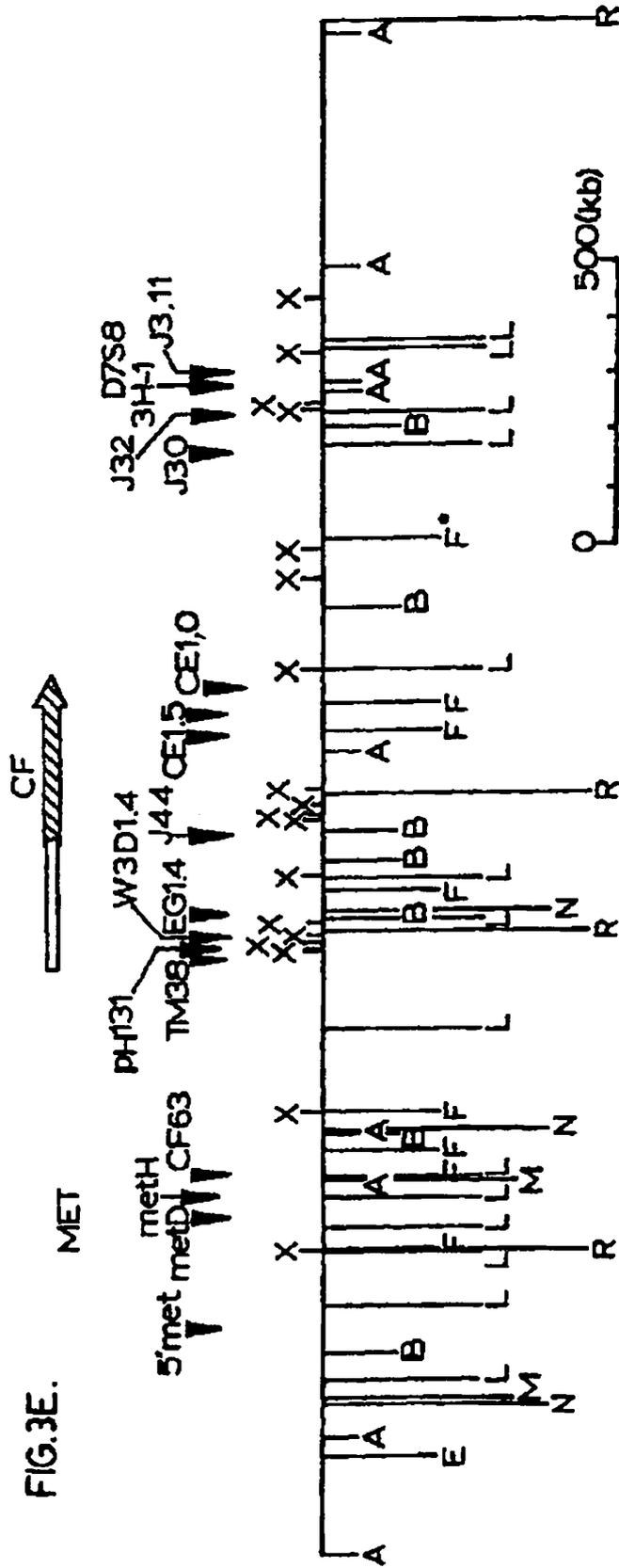


FIG. 4

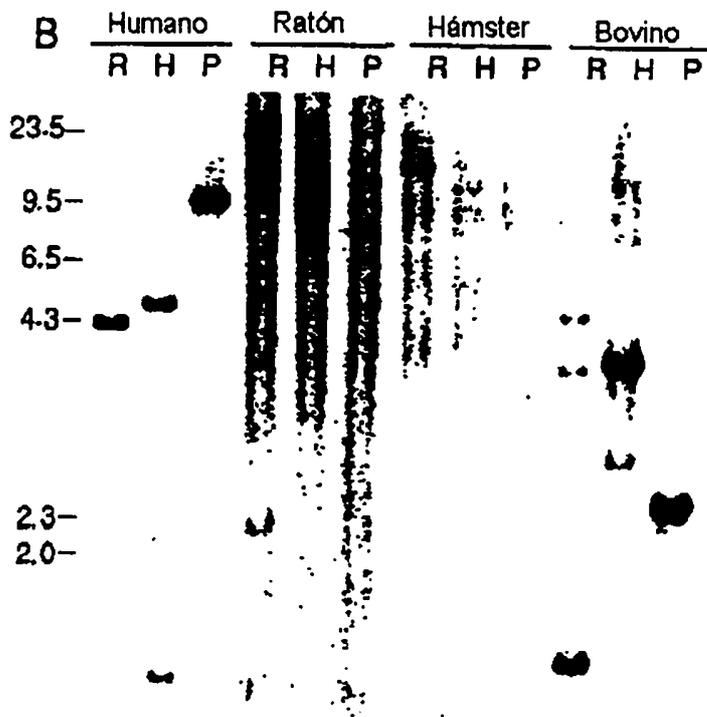
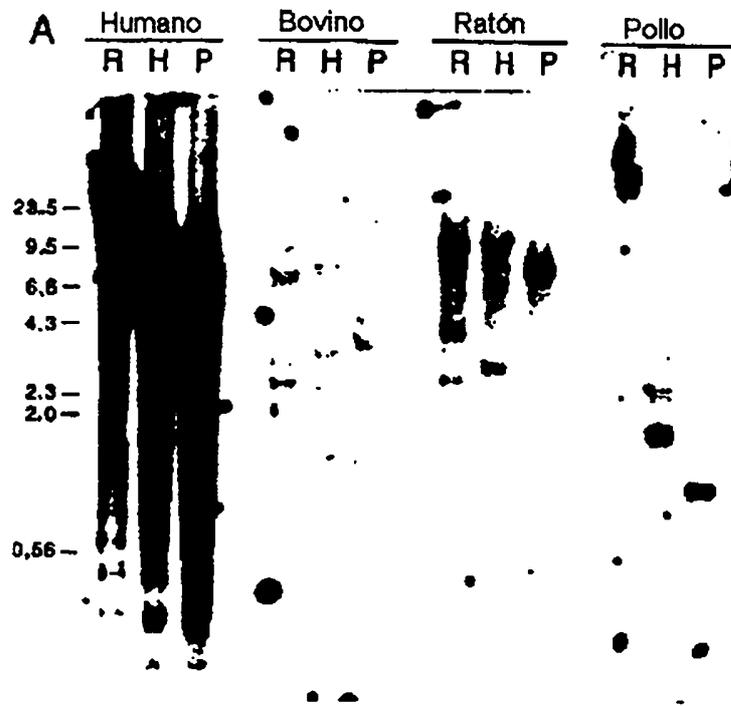
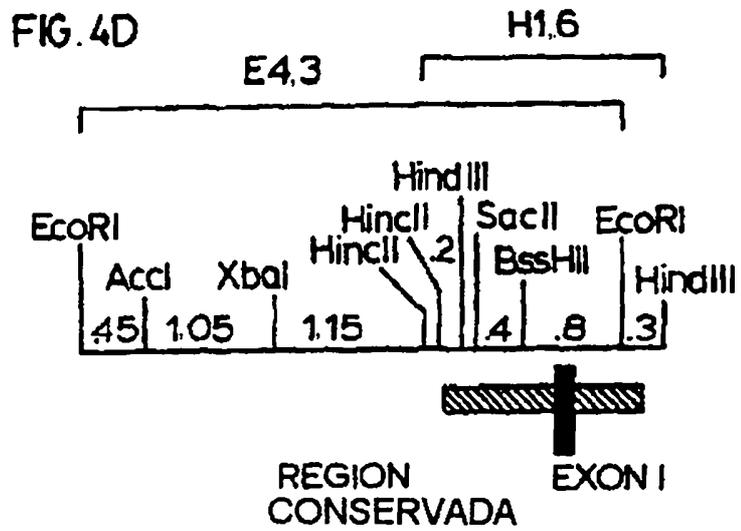
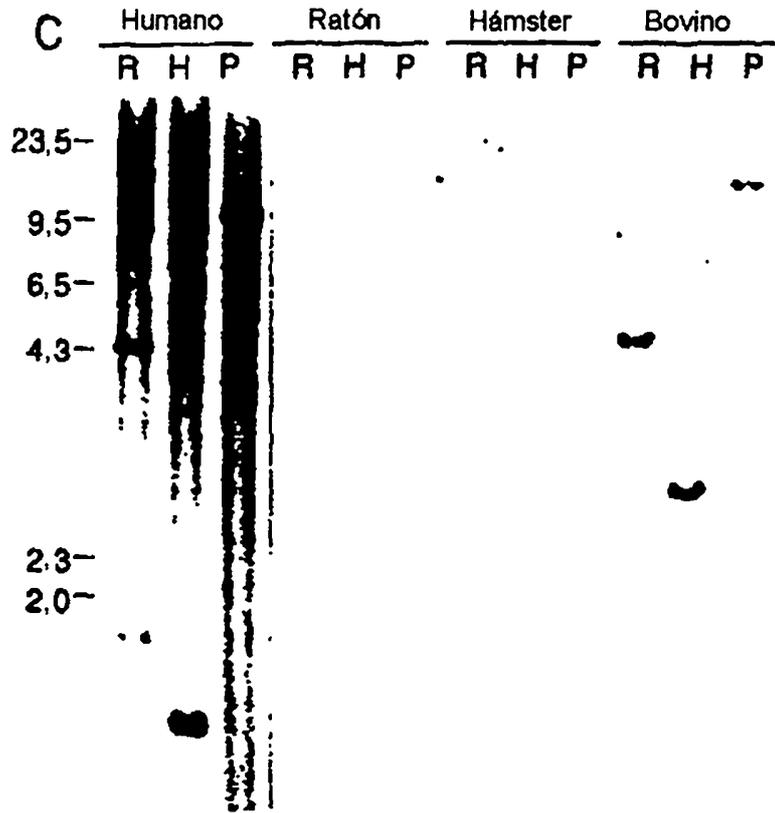


FIG. 4



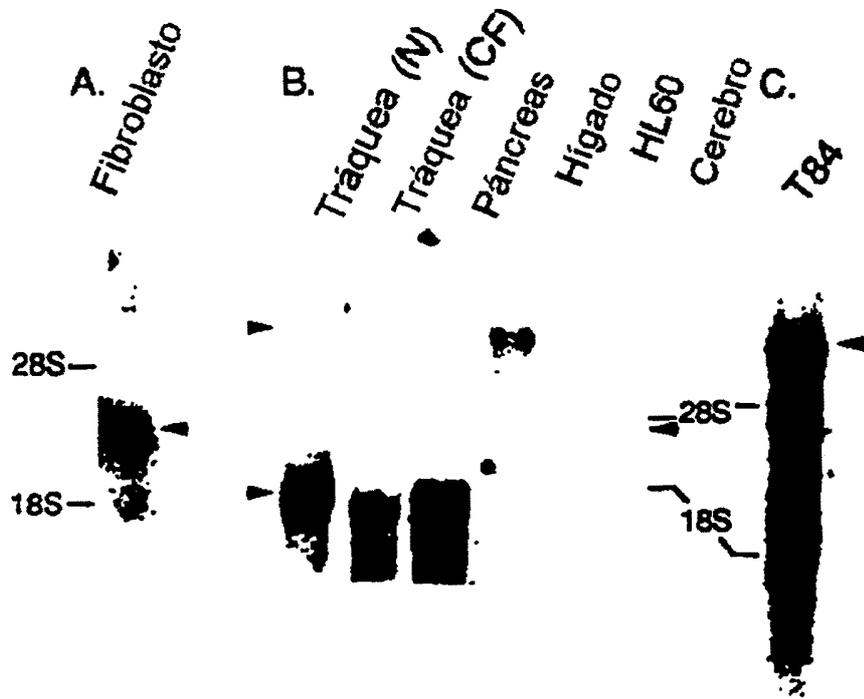


FIG. 6

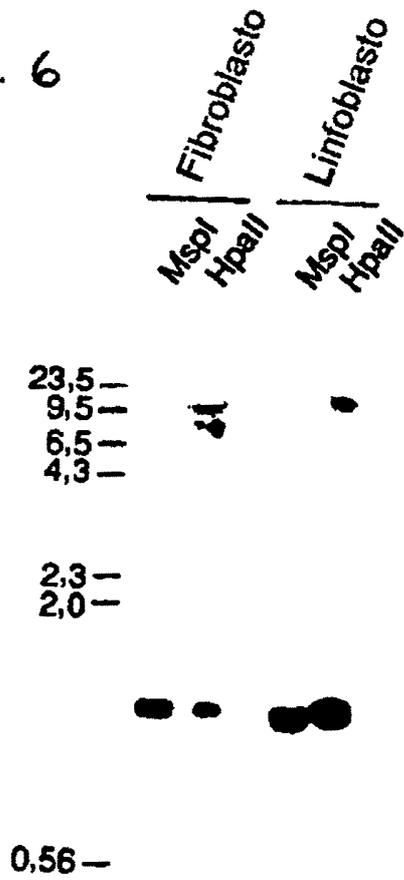


FIG.7.

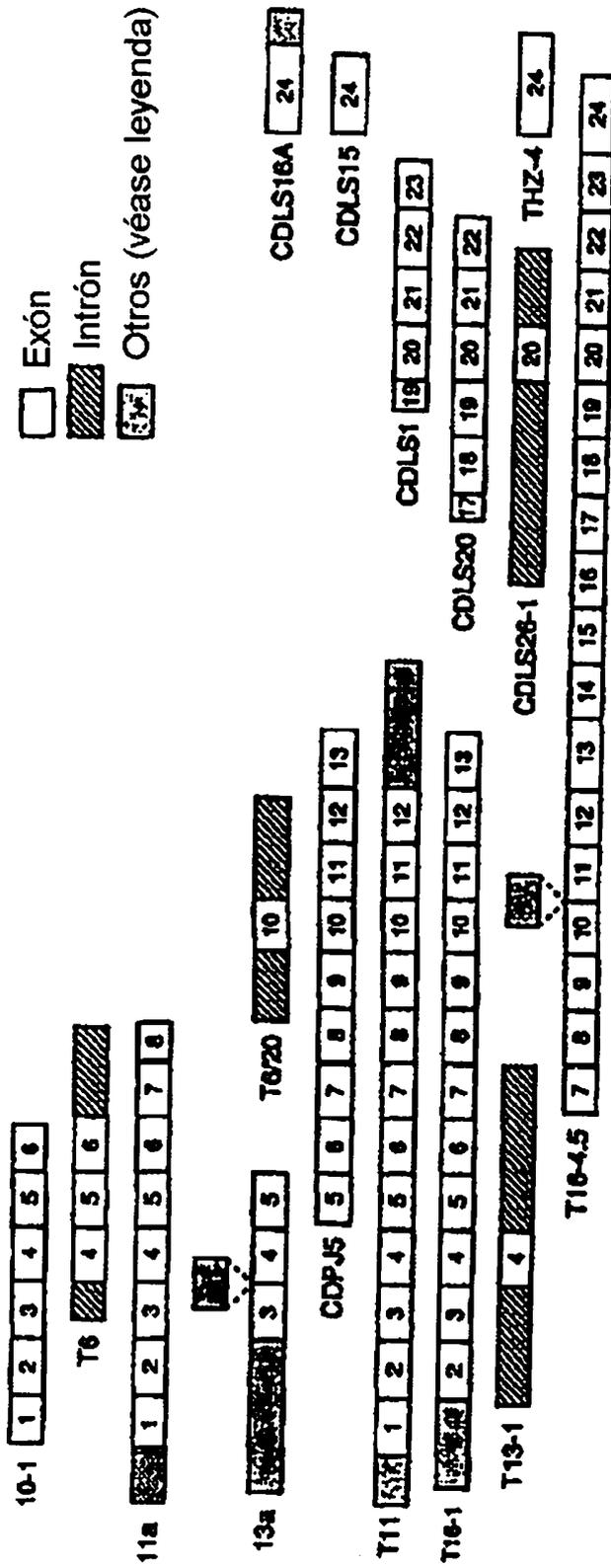
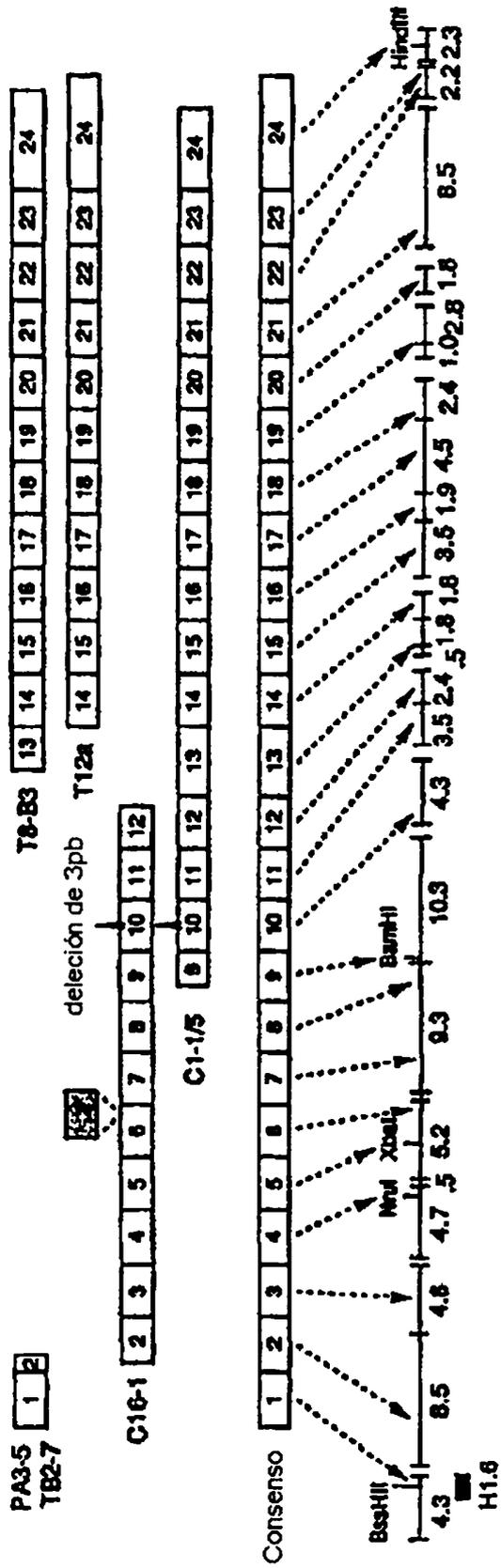


FIG 7. (cont.)



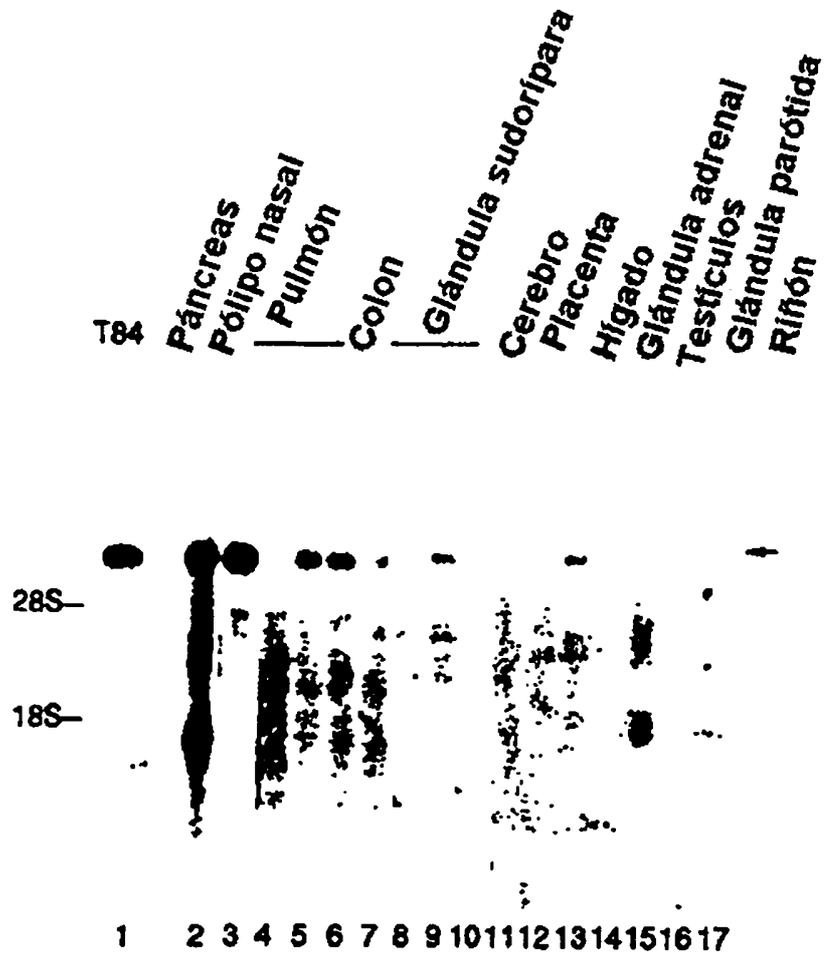


Fig. 8

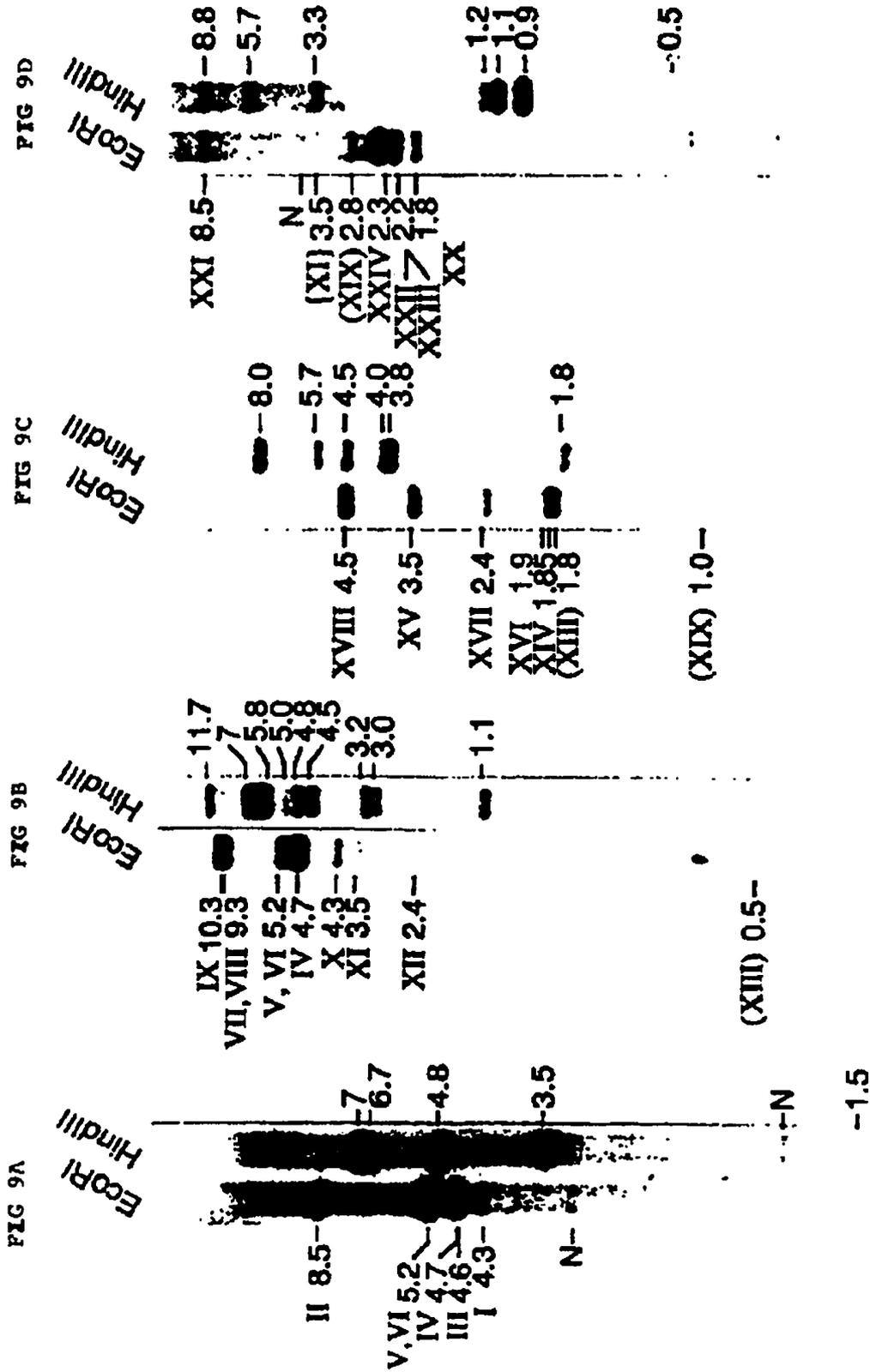


FIG. 10A

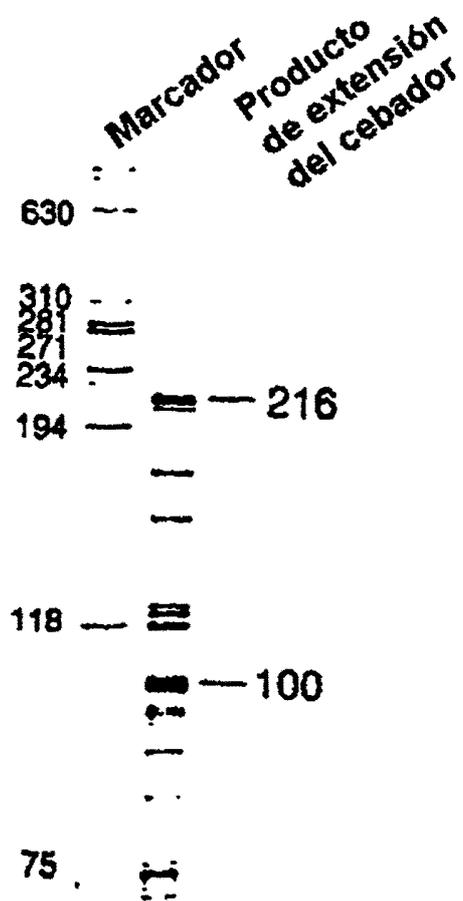


FIG. 10 B

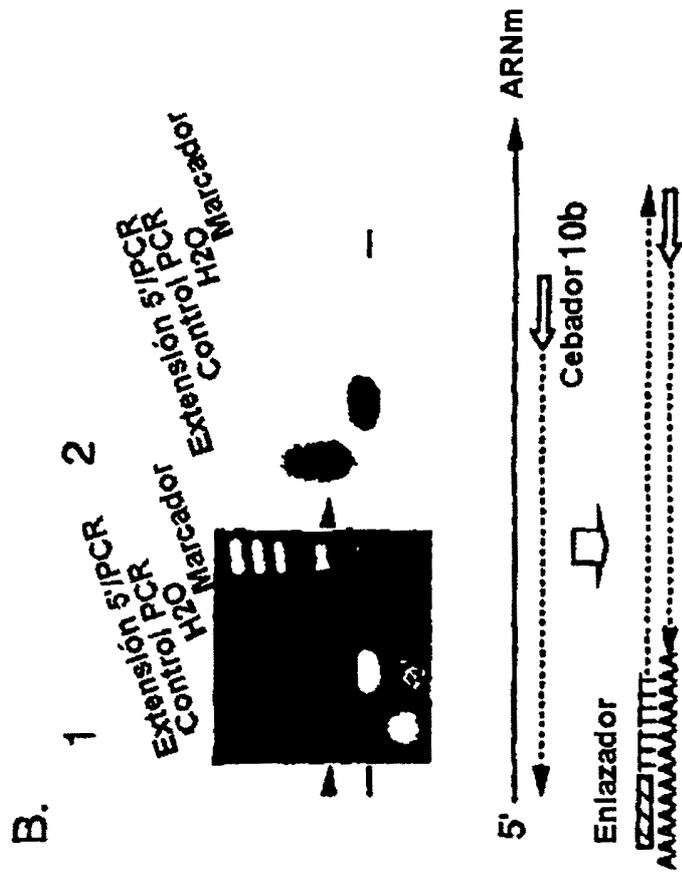


Fig. 10c



FIG.11.

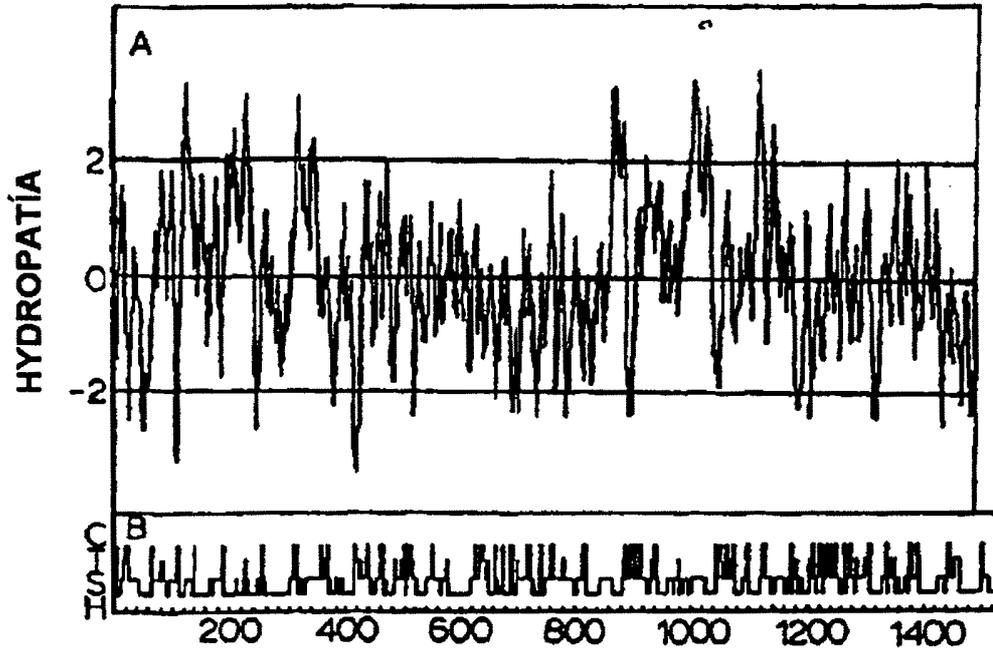


FIG. 12

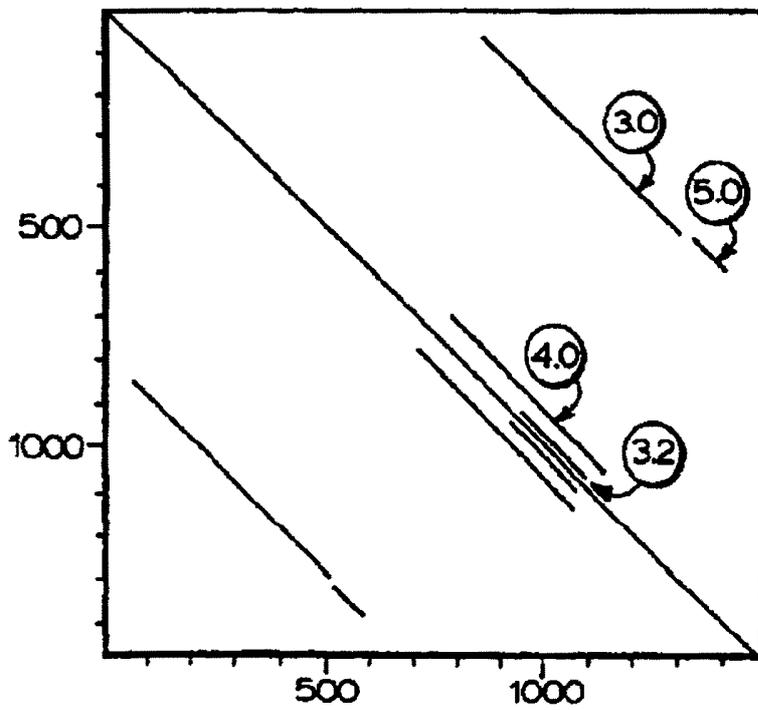


FIG. 13

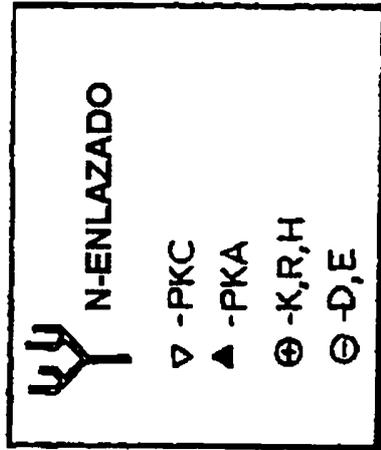
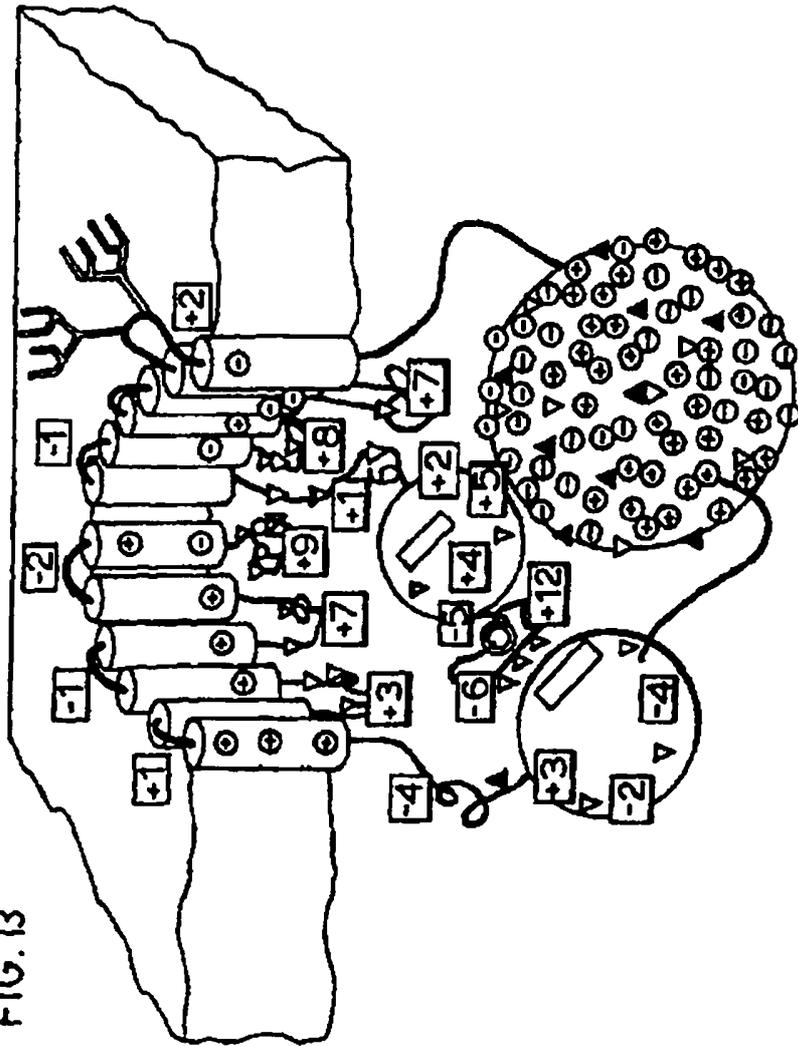
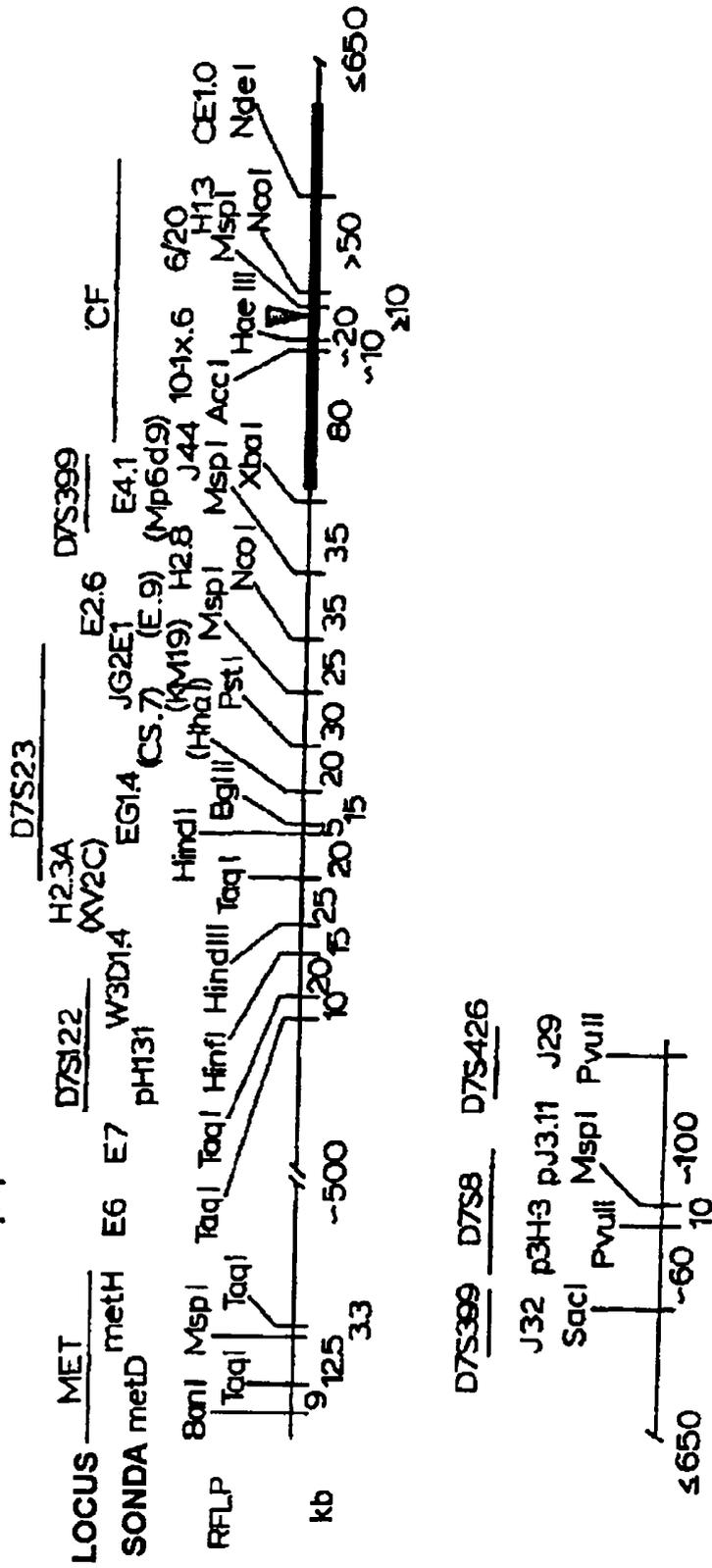
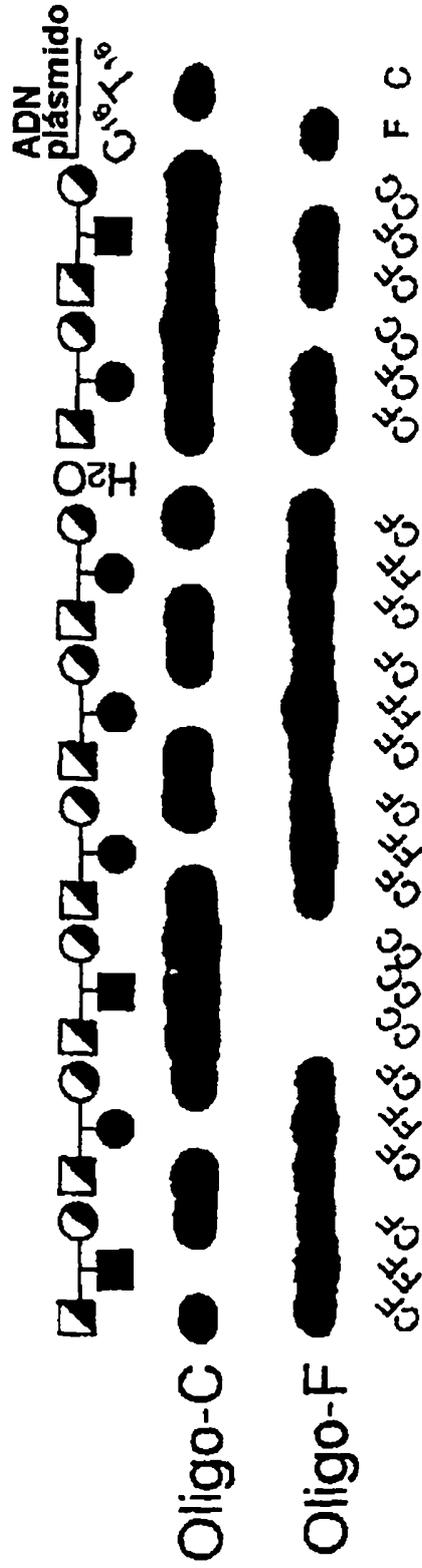


FIG.14.





N: L E N I I F G V
 AAA GAA AAT ATC ATC TTT GGT GTT

CF: L E N I I G V
 AAA GAA AAT ATC AT- --T GGT GTT

Oligo-C: 3' CTTTATAGTAGAACCAC 5'

Oligo-F: 3' TTCTTATAGTA---ACCACAA 5'

Fig. 15

TI G.16.

ISFCSQFSWIMPGTIK-ENIIFGVSYD
 DSITLQWRKAFGVIPQKVFIFSGTFR
 IGVVSQEPVLFATTI-AENIRYGRNV
 LGIVSQEPILFDCSI-AENIRYGDNSR
 IGVVSQEPVLFATTI-AENIRYGREV
 LGEVVSQEPILFDCSI-AENIRYGDNSR
 IGVVSQEPVLSFTTI-AENIRYGRGNV
 LGIVSQEPILFDCSI-AENIRYGDNSR
 IGVVSQEPILFDCSI-KNNIKYSLYSL
 FSIVSQEPMLFNMSI-YENIKFGREDA
 ITVVEQRCTLFNDTL-RKNILLGSTDS
 ISWVEQKPLLFNGTI-RDNILTYGLQDE
 VGVVLQDNVLLNRSI-IDNISLAPGMS
 RCAYVQQDDLF IGLIAREHLIFQAMVR
 MSFVFOHYALFKHMTVYENISFGLRLR
 YLSQQQTPPFATPVWHYLLTLHQHDKTR
 VGMVFOKPTFPFMSI-YDNIAFGVRLF
 GIMVFOHFNLSHMTVLENVMEAPIQV
 VGMVFOSYALYPHLSVAENMSFGLKPA
 ISMIFQDPMTSLNPNYMRVGEOLMEVLM
 IQMIFQDPFLASLNPRTIGEIIAEPRLR
 AGIIHQELNLIPOLTIAENIFLGREFV
 ISEDRKRDGLVLGMSVKENMSLTALRY
 TYTGVFTFVRELFAGVPESTRARGYTEG
 IGVVSQEDNLDLEFTVRENLLVYGRYE
 IGMIFQDHHLLMDRTVYDNVAIPLIIA

FSLLGTPVLKDNFKIERGQLLAVAGSTGAGKSTLLMMING
 YTEGGNAILENISFSISPCQRVCLLGRTCGSKSTLLSAFLR
 PSRKEVKILKGLNKVQSGQTVALVGNSSCGGKSTTVQLMQR
 PTRPDIPVLQGLSLEVKKGQTLALVGNSSCGGKSTTVQLLER
 PSRSEVQILKGLNKVKSQTVALVGNSSCGGKSTTVQLMQR
 PTRPNIPVLQGLSLEVKKGQTLALVGNSSCGGKSTTVQLLER
 PSRANIKILKGLNKVKSQTVALVGNSSCGGKSTTVQLLQR
 PTRANVPVLQGLSLEVKKGQTLALVGNSSCGGKSTTVQLLER
 DTRKDVVEIYKDLSTLLKEGKYAFVGESCGGKSTLLKLE
 ISRPNVPIYKNLSFTCDSSKTTAIVGETSGGKSTFMNLLLR
 PSRPEAVLKNVSNFSAGQFTFIVGKSGGKSTLSNLLLR
 PSAPTAFVYKNNMNFEDMFCQOTLGIIGESGKSTLVLLTK
 YKPDSPVILLDNINISIKQGEVIGVGRSGGKSTLILKLIQR
 IPAPRRKHLKXVCGVAYPCCELLAVMSSGAGKSTLLNALAF
 KSLGNLKIILORVSLYVPKFSLLIALLGPSGSGKSSLIRILAG
 ODVAESTRLGPLSGEVRAGRILHLVGNPAGKSTLLARIAG
 FYYGKFHALKXINLD TAKNOVTAFI GPSGCGKSTLLRFTNK
 RRYGGHEVLKGVSLQARAGOVISIIGSSGSGKSTFLRCINF
 KAWGEVVVSKDINID IHEGEFVTVVGPSCGKSTLLRMIAG
 TPDGDVAVNDLNFTRAGETLIVGESGSGKSTQAFALMG
 QPPKTLKAVDGVTLRLYEGETLIVGESGSGKSTFARAIIG
 KAVPGVKALSGAALNVYPCRVMALVGENGAGKSTMMKVLTG
 VDNLCGPGVNDVSTLIRKGEILGVSGLMGAGRTELKMLVLYG
 LTGARGNNLKDVTLLPVLFTCITGVSGGKSTLINDTLF
 KSYGGKIIVNDLSFTLAAGECFGLLGNAGKSTIIRMLIG
 AYLGGRQALQGVTFHMQPGEAMFLTGHSGAGKSTLLKLIIG

CFTTR (N)
 CFTTR (C)
 hmdr1 (N)
 hmdr1 (C)
 mmdr1 (N)
 mmdr1 (C)
 mmdr2 (N)
 mmdr2 (C)
 pfmdr (N)
 pfmdr (C)
 STE6 (N)
 STE6 (C)
 hlyB
 White
 Hbpx
 Btud
 PstB
 hisP
 malK
 oppD
 oppF
 RbsA (N)
 RbsA (C)
 Uvra
 NodI
 FtsE

FIG. 16. (cont.)

CFTR (N)	GEGITLSSGGQRARISLARAVYKDADLYLLDSSPFGYLDVLTEK
CFTR (C)	VDGGCVLSHGKQIMCLARSVLSKAKILLLLDEPSAHLDPVTYQ
hmdr1 (N)	GERGAQLSSGGQKORIAARALVRNPKILLDEATSAALDTESEA
hmdr1 (C)	GDKGTLSSGGQKORIAARALVRNPKILLDEATSAALDTESEK
mmdr1 (N)	GERGAQLSSGGQKORIAARALVRNPKILLDEATSAALDTESEA
mmdr1 (C)	GDKGTLSSGGQKORIAARALVRNPKILLDEATSAALDTESEK
mmdr2 (N)	GDRGAQLSSGGQKORIAARALVRNPKILLDEATSAALDTESEA
mmdr2 (C)	GDKGTLSSGGQKORIAARALVRNPKILLDEATSAALDTESEK
pfmdr (N)	GSNASKLSSGGQKORISARAIMRNPKILLDEATSSLDNKSEY
pfmdr (C)	PYGKS -LSSGGQKORIAARALLREPKILLDEATSSLDNSSEK
STE6 (N)	GTGGVTLLSSGGQQRVAIARAFIRDTPILFLDEAVSALDIVERN
STE6 (C)	RIDTTLSSGGQQRVAIARALVRNPKILLDEATSAALDVSSS
hlyB	GEQGAGLSSGGQKORIAARALVRNPKILLDEATSAALDYASEH
White	PGRVKGLSSGGERKRLAFASEALTDPPLLICDEPTSGLDSTAH
MbpX	FYPAQLSSGGQKORVALARSLAIQPDLLL - DEPFGLDGLRR
StuD	GRSTNQLSSGGENQRVLAQVLAQITLLLLDEPMNSLDVAQOSA
PatB	HQSGYLSLSSGGQQRLLCIARGIAIRPEVLLLEPCSAALDPISGTG
hisp	GKYPVHLSGGQQRVSIARALAMEP DVLLFDEPTSAALDPELVG
malK	DRKPKALSSGGQQRVAIGRTLVAEP SVFLLDEPLSNLDAALRV
oppD	KMYPHEFSGGMRQVRMIAMALLCRPKLLIADEPTTALDVTVOA
oppF	NRYPHFSGGQQRIGIARALILEPKLIICDDAVSALDVSIOA
RbsA (N)	DKLVGDLSIGDQQMVEIAKVLSEKVIIMDEPTCALIDTETE
RbsA (C)	EQALGLSSGGNQQKVAIARGLMTRPKVLLLEPTPGVDVVGAKK
UVIA	GQSATLSSGGEAQRVKLARELSKRGLYILDEPTTGLHFADIQQ
NodI	NTRVADLSSGGMKRRLLTAGALINDPQLLLLEPTTGLDHPHARH
Ftse	KNFP IQLSGGEQQRVGIARAVVNKPAVLLADEPTGNLDDALSE

FIG. 17

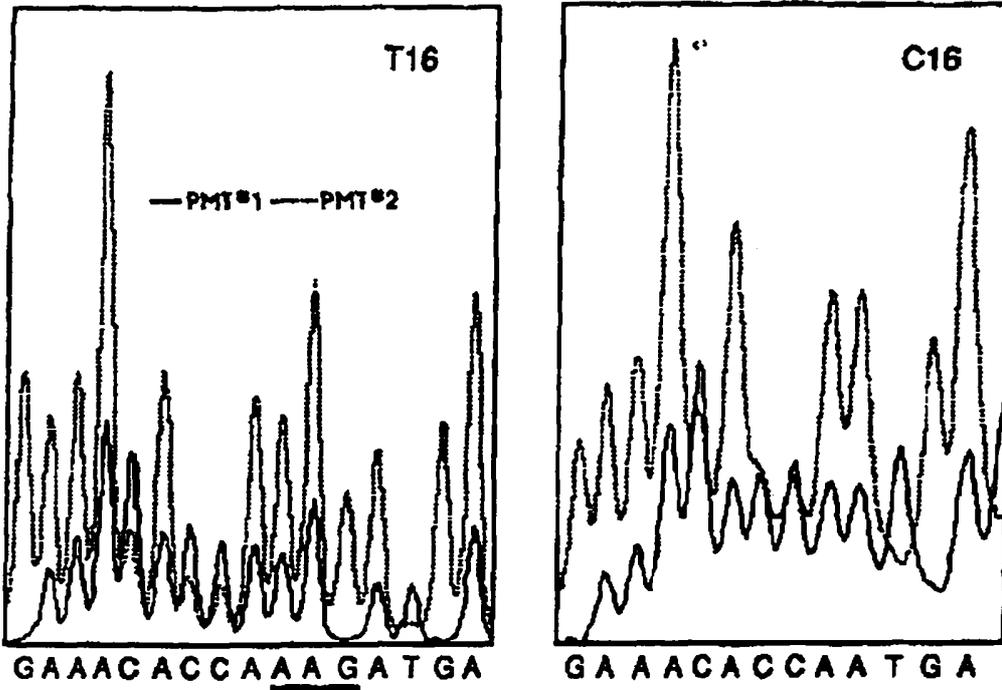


FIG. 18 A

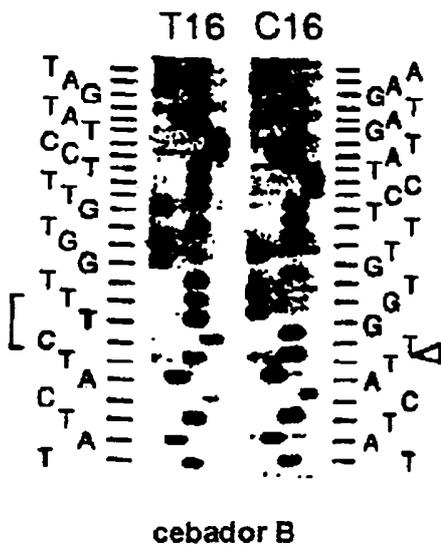


FIG. 18 B

