



Patent- og  
Varemærkestyrelsen

(12) Oversættelse af  
europæisk patentskrift

(51) Int.Cl.<sup>7</sup>: C 12 N 15/12      A 01 K 67/027      A 61 K 48/00      C 07 K 13/00      C 07 K 14/47  
                  C 12 P 21/02      C 12 P 21/08      C 12 Q 1/68      G 01 N 33/53      G 01 N 33/68

(45) Oversættelsen bekendtgjort den: 2004-03-22

(80) Dato for Den Europæiske Patentmyndigheds  
bekendtgørelse om meddelelse af patentet: 2003-11-05

(86) Europæisk ansøgning nr.: 90912428.1

(86) Europæisk indleveringsdag: 1990-08-20

(87) Den europæiske ansøgnings publiceringsdag: 1992-06-10

(86) International ansøgning nr.: CA1990000267

(87) Internationalt publikationsnr.: WO1991002796

(30) Prioritet: 1989-08-22 US 396894      1989-08-24 US 399945      1989-08-31 US 401609

(84) Designerede stater: AT BE CH DE DK ES FR GB IT LI LU NL SE

(73) Patenthaver: HSC Research Development Corporation, 88 Elm Street, Toronto, Ontario, M5G 1X8, Canada  
THE BOARD OF REGENTS ACTING FOR AND ON BEHALF OF THE UNIVERSITY OF MICHIGAN, 475 E. Jefferson,  
Room 2354, Ann Arbor, MI 48109-1248, USA

(72) Opfinder: TSUI, Lap-Chee, 94 Willowridge Road, Toronto, Ontario M9R 3Z4, Canada  
RIORDAN, John, R., 137 Bedford Road, Toronto, Ontario M5R 2K7, Canada  
COLLINS, Francis, S., 4340 Ceio Church Road, Ann Arbor, MI 48103, USA  
ROMMENS, Johanna, M., 199 Bogert Avenue, Willowdale, Ontario M2N 1L1, Canada  
IANNUZZI, Michael, C., 2073 Chaucer Street, Ann Arbor, MI 48103, USA  
KEREM, Bat-Sheva, 140 Elm Ridge Drive, Apt. 817, Toronto, Ontario M6B 1B1, Canada  
DRUMM, Mitchell, L., 2431 Pinecrest Street, Ann Arbor, MI 48103, USA  
BUCHWALD, Manual, 15 Bearbourne Avenue, Toronto, Ontario M4K 1M6, Canada

(74) Fuldmægtig i Danmark: Zacco Denmark A/S, Hans Bekkevolds Allé 7, 2900 Hellerup, Danmark

(54) Benævnelse: Cystisk fibrose gen

(56) Fremdragne publikationer:  
EP-A- 0 288 299  
Science, Volume 245, No. 4922, 8 September 1989, (Washingt-  
on, D.C., US) J.R. RIORDEN et al.: "Identification of the Cy-  
stic Fibrosis Gene: Cloning and Characterization of Comple-  
mentary DNA", pages 1066-1073 see the whole article  
Science, Volume 245, No. 4922, 8 September 1989, (Washingt-  
on, D.C., US) B.S. KEREM et al.: "Identification of the Cystic  
Fibrosis Gene: Genetic Analysis", pages 1073-1080 see the  
whole article



Den foreliggende opfindelse angår generelt cystisk fibrose (CF) gen og mere specifikt identificeringen, isolationen og kloningen af den DNA-sekvens, der svarer til de normale og mutante CF gener, såvel som deres transcripter og genprodukter. Den foreliggende opfindelse angår også fremgangsmåder til screening for og påvisning af CF bærere, CF diagnose, prænatal CF screening og diagnose, og genterapi, som benytter rekombinante teknologier, og lægemiddelterapi, som anvender informationen afledt fra DNA'en, proteinet og proteinets metaboliske funktion.

Cystisk fibrose (CF) er den mest almindelige autosomalt recessive genetiske lidelse i den kaukasiske population. Den berører ca. 1 ud af 2000 levendefødte i Nordamerika [Boat et al, The Metabolic Basis of Inherited Disease, 6th ed, pp. 2649 - 2680, McGraw Hill, NY (1989)]. Omtrentligt 1 ud af 20 personer er bærer af sygdommen.

Selv om sygdommen først blev beskrevet sidst i 1930'erne, er den basale defekt stadig ukendt. De væsentlige symptomer på cystisk fibrose indbefatter kronisk pulmonal sygdom, pancreatisk eksokrin insufficiens og forhøjede sved-elektrolyt-niveauer. Symptomerne er konsistente med, at cystisk fibrose er en eksokrin lidelse. Selv om nylige fremskridt er blevet gjort i analysen af ion-transport tværs gennem den apicale membran af epitelet i celler fra CF patienter, er det ikke klart, at den unormale regulering af chlorid-kanaler repræsenterer den primære defekt ved sygdommen. Givet manglen på forståelse af den molekylære mekanisme for sygdommen er der derfor blevet gjort et alternativt approach i et forsøg på at forstå beskaffenheden af den molekylære defekt via direkte kloning af det ansvarlige gen på basis af dets kromosomale lokalisering.

Der er imidlertid ikke nogen klar fænotype, som styrer et approach til den eksakte beskaffenhed af den genetiske basis for sygdommen, eller som tillader en identifikation af cystisk fibrose genet. Beskaffenheden af CF defekten i relation til de populationsgenetiske data har ikke været helt tydelig. Både sygdommens prævalens og den kliniske heterogenitet er blevet forklaret ved adskillige forskellige mekanismer: høj mutationsgrad, heterozygot fordel, genetisk drift, multiple loci og reproduktiv kompensation.

Mange af hypoteserne kan ikke testes på grund af manglen på viden om den basale defekt. Derfor har alternative approaches for bestemmelsen og karakteriseringen af CF genet fokuseret på et forsøg på at identificere lokaliseringen af genet ved hjælp af genetisk analyse.

5

Bindingsanalyse af CF genet til antigene markører og proteinmarkører blev forsøgt i 1950'erne, men ingen positive resultater blev opnået (Steinberg et al Am. J. Hum. Genet. 8: 162 - 176, (1956); Steinberg og Morton Am. J. Hum. Genet. 8: 177 - 189, (1956); Goodchild et al J. Med. Genet. 7: 417 - 419, 1976.

10

Mere for nylig er det blevet muligt at anvende RFLP'er til at fremme bindingsanalyse. Den første binding af en RFLP markør til CF genet blev beskrevet i 1985 (Tsui et al. Science 230: 1054 - 1057, 1985], hvor binding blev fundet mellem CF genet og en ukarakteriseret markør DOCRI-917. Associationen blev fundet i en analyse af 39 familier med berørte CF børn. Dette viste, at selv om den kromosomale lokalisering ikke var blevet etableret, var lokaliseringen af sygdomsgenet blevet indsnævret til omtrentligt 1 % af det humane genom, eller omtrentligt 30 millioner nucleotidbasepar.

15

Den kromosomale lokalisering af DOCRI-917 proben blev etableret under anvendelse af gnaver-humane hybride cellelinjer, der indeholdt forskellige humane kromosom-komplementer. Det blev vist, at DOCRI-917 (og derfor CF genet) map'er til humant kromosom 7.

20

Yderligere fysiske og genetiske bindingsstudier blev udført i et forsøg på at pinpointe lokaliseringen af CF genet. Zengerling et al [Am. J. Hum. Genet. 40: 228 - 236 (1987)] beskriver anvendelsen af human-muse somatiske cellehybrider med henblik på at opnå et mere detaljeret fysisk forhold mellem CF genet og de markører, der er kendt for at være forbundet dermed. Denne publikation viser, at CF genet kan være tildelt til enten den distale region af bånd q22 eller den proksimale region af bånd q31 på kromosom 7.

25

30

Rommens et al [Am. J. Hum. Genet. 43: 645 - 663, (1988)] giver en detaljeret diskussion af isolationen af mange nye 7q31 prober. Det beskrevne approach førte til isolationen af to nye prober, D7S122 og D7S340, der er tæt på hinanden. Pulset felt gel elektroforese mapping viser, at disse to RFLP markører er mellem to markører kendt for at flankere CF genet, MET [White,

35

R., Woodward S., Leppert M., et al. Nature 318: 382 - 384, (1985)] og D7S8 [Waitright, B. J., Scambler, P. J. og J. Schmidtke, Nature 318: 384 - 385 (1985)], derfor i CF gen regionen. Opdagelsen af disse markører tilvejebringer et udgangspunkt for kromosom walking og jumping.

5

Estivill et al, [Nature 326: 840 - 845 (1987)] beskriver, at et kandidat cDNA gen var lokaliseret og delvist karakteriseret. Dette beskriver imidlertid ikke den korrekte lokalisering af CF genet. Referencen beskriver et kandidat cDNA gen nedstrøms for en CpG ø, der er undermethylerede GC nucleotid-rige regioner opstrøms for mange vertebrate gener. Den kromosomale lokalisering af kandidat locus'et er identificeret som XV2C regionen. Denne region er beskrevet i Europæisk Patentansøgning 88303645.1. Imidlertid indbefatter den faktiske region ikke CF genet.

10

15

En større vanskelighed i at identificere CF genet har været manglen på cytologisk påviselige kromosom-omflytninger eller -deletioner, der i høj grad fremmede alle tidligere succeser i kloningen af humane sygdomsgener ved hjælp af viden om map position.

20

Sådanne omflytninger og deletioner kunne observeres cytologisk, og som en følge deraf kunne en fysisk lokalisering på et særligt kromosom korreleres med den givne sygdom. Yderligere kunne denne cytologiske lokalisering korreleres med en molekylær lokalisering baseret på kendte forhold mellem offentligt tilgængelige DNA prober og cytologisk synlige ændringer i kromosomerne. Viden om den molekylære lokalisering af genet for en given sygdom ville muliggøre kloning og sekventering af det pågældende gen ved hjælp af rutine procedurer, især når genproduktet er kendt, og kloningssucces kan bekræftes ved hjælp af immunoassay af ekspressionsprodukter af de klonede gener.

25

30

I modsætning hertil var hverken den cytologiske lokalisering eller genproduktet af genet for cystisk fibrose kendt inden for det tekniske område. Med den nylige identifikation af MET og D7S8, markører som flankerede CF genet, men ikke pinpointede dets molekylære lokalisering, udviklede de foreliggende opfindere forskellige hidtil ukendte genkloningsstrategier for at nærme sig CF genet ifølge den foreliggende opfindelse. De benyttede fremgangsmåder i disse strategier indbefatter kromosom jumping fra de flankerende markører, kloning af DNA fragmenter fra en defineret fysisk region med anvendelsen af

35

pulset felt gel elektroforese, en kombination af somatisk cellehybrid og molekylære kloningsteknikker designet til at isolere DNA fragmenter fra undermethylerede CpG øer nær CF, kromosom-mikrodissektion og -kloning, og mætning kloning af et stort antal af DNA markører fra 7q31 regionen. Ved hjælp af disse hidtil ukendte strategier var de foreliggende opfindere i stand til at identificere det gen, der er ansvarligt for cystisk fibrose, hvor den kendte teknik var uvis eller endog i ét tilfælde, forkert.

Anvendelsen af disse genetiske og molekylære kloningsstrategier har muliggjort isolationen og cDNA kloningen af cystisk fibrose genet på basis af dets kromosomale lokalisering, uden fordelen ved genomiske omflytninger til at udpege vejen. Identifikationen af de normale og mutante former af CF genet og genprodukter har muliggjort udviklingen af screening og diagnostiske tests for CF under anvendelse af nucleinsyreprober og antistoffer mod genproduktet. Via interaktion med det defekte genprodukt og den pathway, hvor dette genprodukt er involveret, er terapi via normal genprodukt supplering og gen manipulation og tildeling nu mulig.

Det gen, der er involveret i sygdomsprocessen for cystisk fibrose, herefter "CF genet", og dets funktionelle ækvivalenter, er blevet identificeret, isoleret og cDNA klonet, og dets transcripter og genprodukter identificeret og sekventeret. En tre basepar deletion, der fører til udeladelsen af en phenylalanin rest i genproduktet, er blevet bestemt til at korrespondere til mutationerne af CF genet i ca. 70 % af de patienter, der er ramt af CF, idet forskellige mutationer er involveret i de fleste, hvis ikke alle de resterende tilfælde.

Med identifikationen og sekventeringen af genet og dets genprodukt kan nucleinsyreprober og antistoffer rejst mod genproduktet anvendes i en række af hybridiseringsassays og immunologiske assays til at screene for og påvise tilstedeværelsen af enten et normalt eller et defekt CF gen eller genprodukt. Assay kits for en sådan screening og diagnose kan også være tilvejebragt.

Patientterapi via supplering med det normale genprodukt, hvis produktion kan amplificeres under anvendelse af genetiske og rekombinante teknikker, eller dets funktionelle ækvivalent, er nu også mulig. Korrektur eller modifikation af det defekte genprodukt via lægemiddelbehandlingsmidler kan nu være mulig. Derudover kan cystisk fibrose kureres eller kontrolleres via genterapi ved at

korrigere gen-defekten *in situ* eller ved at anvende rekombinante eller andre vehikler til at tildele en DNA sekvens, der er i stand til ekspression af det normale genprodukt til patientens celler.

5 Ifølge et aspekt ifølge den foreliggende opfindelse er der tilvejebragt et oprenset DNA molekyle, der omfatter et cystisk fibrose gen, idet nævnte gen omfatter en DNA sekvens udvalgt fra gruppen bestående af:

10 (a) DNA sekvenser, der svarer til DNA sekvensen som i figur 1 fra aminosyrerestposition 1 til position 1480 eller en polymorf form af nævnte DNA sekvens, hvor nucleotid-substitution i nævnte DNA sekvens ikke påvirker den essentielle funktion af polypeptidet kodet derved;

15 (b) DNA sekvenser, der koder et normal cystisk fibrose transmembrane konduktans regulator (CFTR) polypeptid med sekvensen ifølge figur 1 fra aminosyrerestposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyre-substitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf, eller dets hydrofatiske profil eller sekundære eller tertiære struktur;

20 (c) DNA sekvenser, som svarer til et fragment af DNA sekvensen i figur 1, herunder mindst 16 sekventielle nucleotider mellem aminosyrerestposition 134 og 4573, og som kan screene for og påvise tilstedeværelsen af enten et normalt eller defekt cystisk fibrose gen eller genprodukt;

25 (d) DNA sekvenser, som omfatter mindst 16 nucleotider og koder et fragment af aminosyresekvensen i figur 1, som kan screene for og påvise tilstedeværelsen eller genproduktet; og

30 (e) DNA sekvenser, der koder en epitop kodet af mindst 18 sekventielle nucleotider i sekvensen i DNA sekvensen i figur 1 mellem aminosyrerestposition 1 og 1480.

DNA'en kan for eksempel være cDNA.

35

Ifølge et andet aspekt ifølge den foreliggende opfindelse, der er tilvejebragt en oprenset DNA molekyle af (a), (b) eller (c) som defineret ovenfor, yderligere

indeholdende mindst én mutation i den DNA sekvens, der, hvis den udtrykkes i celler i det menneskelige legeme, er associeret med ændret cellefunktion, som korrelerer med cystisk fibrose, hvor nævnte mutation indbefatter en deletion af tre nucleotider, der koder phenylalanin ved aminosyrerestposition 508.

5

Ifølge et andet aspekt ifølge den foreliggende opfindelse, der er tilvejebragt oprenset nucleinsyre probe omfattende en DNA eller RNA nucleotid-sekvens svarende til sekvensen ifølge egenskaber (c), (d) eller (e) som defineret ovenfor.

10

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en rekombinant kloningsvektor, som omfatter et DNA molekyle ifølge den foreliggende opfindelse. Vektoren ifølge et aspekt ifølge den foreliggende opfindelse er operativt forbundet til en ekspressionskontrolsekvens i det rekombinante DNA molekyle, således at det normale CFTR protein kan udtrykkes, eller alternativt kan med den anden selekterede mutante DNA sekvens det mutante CFTR polypeptid udtrykkes. Ekspressionskontrolsekvensen er udvalgt fra gruppen bestående af sekvenser, der kontrollerer ekspressionen af gener i prokaryote eller eukaryote celler og deres virusser og kombinationer deraf.

15

I et andet aspekt omfatter den foreliggende opfindelse en non-human vært transformeret med en vektor ifølge den foreliggende opfindelse.

20

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en fremgangsmåde til fremstilling af et normalt eller mutant cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid, hvilken fremgangsmåde omfatter trinene med:

25

(a) dyrkning af en værtscelle transficeret med vektoren ifølge den foreliggende opfindelse i et medium og under betingelser, der er favorable for ekspression af normalt eller mutant CFTR polypeptid; og

(b) isolering af det udtrykte normale eller mutante CFTR polypeptid.

30

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt et oprenset DNA molekyle, der omfatter en DNA sekvens, som koder et mutant

35



5 cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid, der har sekvensen ifølge figur 1 for aminosyrerestposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid, der har sekvensen ifølge figur 1, ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur, yderligere er kendetegnet ved en deletion af tre nucleotider, der koder phenylalanin ved aminosyrerestposition 508.

10 Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt et oprenset cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid kendetegnet ved at have en peptidmolekylvægt på omtrentligt 170.000 dalton udvalgt fra gruppen bestående af:

15 (a) et normalt CFTR polypeptid, der har celle transmembran ion konduktans påvirkende aktivitet og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte normal CFTR polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 påvirker ikke den essentielle funktion deraf, eller dets hydropatiske profil eller sekundære eller tertiære struktur; og

20 (b) et mutant CFTR polypeptid med cystisk fibrose aktivitet i humane celler og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyrerestposition 508.

25 Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt et polypeptid kodet af ekspresion af en DNA sekvens ifølge den foreliggende opfindelse, idet nævnte polypeptid udviser den immunologiske eller biologisk aktivitet af normalt eller mutant CFTR polypeptid.

30 Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en fremgangsmåde til isolering af det normale eller mutante CFTR protein ifølge den foreliggende opfindelse fra celler, der indeholder nævnte protein, hvilken fremgangsmåde omfatter trinene med:

35 (a) opløsning af protein fra en selekteret cellemembran, hvor enten nævnte normale eller mutante CFTR protein er udtrykt, med henblik på at tilvejebringe en opløsning af nævnte CFTR protein:

(b) adskillelse af nævnte CFTR protein fra nævnte opløsning ved at kontakte nævnte opløsning med antistoffer mod nævnte normale eller mutante CFTR protein, idet nævnte antistoffer er immobiliseret på et substrat;

5 (c) skylning af nævnte substrat med henblik på at fjerne protein, der ikke klæbede til nævnte antistoffer;

(d) frigivelse af nævnte CFTR protein fra nævnte antistoffer for derved at isolere nævnte CFTR protein, og

10

(e) oprensning af nævnte CFTR protein for at fjerne eventuelt resterende andet mammalt protein.

15 Ifølge et andet aspekt ifølge den foreliggende opfindelse omfatter en fremgangsmåde til at screene et individ med henblik på bestemme, om nævnte individ er bærer af cystisk fibrose (CF) bærer eller en CF patient:

20 tilvejebringelse af et assay til påvisning i en biologisk prøve, der skal screenes, af tilstedeværelsen af mindst et medlem fra gruppen bestående af et normalt CF gen, normalt CF gen produkter, et mutant CF gen, mutant CF genprodukter og blandinger deraf, et normalt CFTR polypeptid med celle transmembran ion konduktans påvirkende aktivitet og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte normale CFTR polypeptid, hvor aminosyresubstitution i de variable regioner af

25 nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur, et mutant CFTR polypeptid med cystisk fibrose aktivitet i humane celler og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyrerestposition 508, hvor assayet er

30

(i) et assay, der involverer hybridisering, eller

(ii) et immunologisk assay,

35 og/eller er opnået ved at anvende af én eller flere hybridiseringer under anvendelse af specifikke oligonucleotider, direkte DNA sekventering, restriktionszymer, skelnen på basis af elektroforetisk mobilitet i geler med

denaturerende middel, RNase beskyttelse, kemisk spaltning eller den ligase-medierede påvisningsprocedure.

5 Hybridiseringsassayet kan indbefatte anvendelsen af ét eller flere nucleotider ifølge den foreliggende opfindelse, hvor der ikke er to nucleotider, der er ens, hvilke nucleotider kan være mærket ved hjælp af et hvilket som helst af traditionelt middel.

10 Hvor assay-fremgangsmåden er immunologisk, og den biologiske prøve indbefatter et normalt cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid fra individet og/eller et mutant CFTR polypeptid fra individet: nævnte normale CFTR polypeptid med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med  
15 sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur; og nævnte mutante CFTR polypeptid med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyrerestposition 508.

20 Den immunologiske assay-fremgangsmåde kan indbefatte antistoffer specifikke for et normalt og/eller mutant CFTR polypeptid. Antistofferne til anvendelse i fremgangsmåden kan være monoklonale.

25 Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt fremgangsmåden for screening en individ ifølge den foreliggende opfindelse er udført på et foster *in utero*.

30 Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en fremgangsmåde til screening for en potentiel cystisk fibrose (CF) bærer eller patient med henblik på at identificere tilstedeværelsen af en identificeret cystisk fibrose mutation i det CF gen, der koder et cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid med sekvensen ifølge figur 1 fra aminosyreposition 1 til position 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af  
35 nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur, idet nævnte fremgangsmåde indbefatter trinene med:

(a) hybridisering af en DNA probe ifølge krav 3 på isoleret genomisk DNA fra nævnte potentielle CF bærer eller nævnte potentielle patient, idet nævnte DNA probe udspænder nævnte cystisk fibrose mutation i nævnte CF gen, hvor nævnte DNA probe er i stand til at påvise nævnte cystisk fibrose mutation; og

5

(b) behandling af nævnte genomiske DNA med henblik på at bestemme tilstedeværelsen eller fraværet af nævnte DNA probe og derved vise, ifølge en forud bestemt måde for hybridisering, tilstedeværelsen eller fraværet af nævnte cystisk fibrose mutation.

10

I et yderligere aspekt omfatter den foreliggende opfindelse en fremgangsmåde til påvisning af cystisk fibrose bærere eller patienter, hvor nævnte fremgangsmåde består i at kontakte et CF gen fra patienten eller bæreren med en restriktionsendonuclease til bestemmelse af tilstedeværelsen eller fraværet af et restriktionsendonuclease site i CF genet og sammenligning af mønsteret for restriktionsendonuclease-sites i CF genet med et normalt CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerest ved position 1 og 1480 i figur 1, idet tilstedeværelsen af sites foruden det normale CF gen eller fraværet af sites i det normale CF er indikativt for et mutant CF gen.

15

20

I et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en fremgangsmåde til påvisning af cystisk fibrose bærere, hvor nævnte fremgangsmåde består af bestemmelse af differential mobilitet af heteroduplex PCR produkter i polyacrylamidgeler som et resultat af insertioner eller deletioner i det mutante CF gen sammenlignet med det normale CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerest ved position 1 og 1480 i figur 1.

25

30

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt et kit til undersøgelse for tilstedeværelsen af normalt CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerestposition 1 og 1480 i figur 1 eller et mutant CF gen med mindst en tre basepar deletion i det normale CF gen, hvilken deletion resulterer i deletionen af en phenylalanin fra aminosyrerestposition 508 i figur 1, ved hjælp af immunoassay, hvilket kit omfatter:

35

(a) et antistof, der specifikt binder til et genprodukt af det normale eller det mutante CF gen;

5

(b) reagensmidler til påvisning af bindingen af antistoffet mod genproduktet; og

(c) idet antistoffet og reagensmidlerne hver er til stede i mængder effektive til at udføre immunoassayet.

10

Ifølge et yderligere aspekt ifølge den foreliggende opfindelse er der tilvejebragt et kit til undersøgelse for tilstedeværelsen af et normalt CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerestposition 1 og 1480 i figur 1 eller et mutant CF gen med mindst en tre basepar deletion i det normale CF gen, hvilken deletion resulterer i deletionen af en phenylalanin fra aminosyrerestposition 508 af figur 1, ved hybridisering, hvilket kit omfatter:

15

(a) en oligonucleotidprobe, som specifikt binder til det normale eller det mutante CF gen;

20

(b) reagensmidler til påvisning af hybridiseringen af oligonucleotidproben til det normale eller det mutante CF gen; og

(c) idet proben og reagensmidlerne hver er til stede i mængder effektive til at udføre hybridiseringsassayet.

25

Ifølge et yderligere aspekt ifølge den foreliggende opfindelse er der tilvejebragt et immunologisk aktivt anti-cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid polyklonalt eller monoklonalt antistof specifikt for et CFTR polypeptid ifølge den foreliggende opfindelse. En hybridom, der producerer det monoklonale antistof ifølge den foreliggende opfindelse, er et yderligere aspekt ifølge den foreliggende opfindelse.

30

I et andet aspekt omfatter den foreliggende opfindelse en sammensætning til behandlingen af cystisk fibrose i en patient, hvilken sammensætning omfatter en terapeutisk effektiv mængde af et oprenset normal CFTR polypeptid ifølge den foreliggende opfindelse. Bæreren kan indbefatte et lunge surfactant protein med henblik på fremme anvendelse af sammensætningen til respiratoriske epiteliale celler.

35

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt en sammensætning til genterapi for cystisk fibrose, hvilken sammensætning omfatter et DNA molekyle ifølge den foreliggende opfindelse og en vehikel til tildeling af nævnte DNA molekyle til en celle af en cystisk fibrose patient.  
5 Vehiklen kan for eksempel være en rekombinant vektor.

Ifølge et andet aspekt ifølge den foreliggende opfindelse er der tilvejebragt et non-humant dyr, der omfatter et heterologt cellesystem, som omfatter en rekombinant kloningsvektor ifølge den foreliggende opfindelse, hvilken vektor inducerer symptomer på cystisk fibrose i nævnte dyr.  
10

Ifølge et andet aspekt ifølge den foreliggende opfindelse er en transgen mus, der udviser symptomer på cystisk fibrose og er opnåelig ved hjælp af de trin, der omfatter indføring af et cystisk fibrose gen, som koder for et mutant cystisk fibrose transmembran konduktans regulator protein, som indbefatter mindst en deletion af phenylalanin ved den position, der svarer til 508 i figur 1, ind i en muse-oocyt eller et muse-foster;  
15

og eventuelt inaktivering af det endogene cystisk fibrose muse gen.  
20

Med henblik på at den foreliggende opfindelse kan forstås mere fuldstændigt, vil der nu blive henvist til de ledsagende tegninger, hvor:

figur 1 er nucleotid-sekvensen for CF genet og aminosyresekvensen for CFTR proteinet;  
25

figur 2 er et restriktionsmap af CF genet og den skematiske strategi anvendt til kromosom walk og jump for genet;

figur 3 er et pulset-felt-gel elektroforese map af den region, der indbefatter og omgiver CF genet;  
30

figur 4A, 4B og 4C viser påvisningen af konserverede nucleotid-sekvenser ved hjælp af tværgående art hybridisering;  
35

figur 4D er et restriktionsmap af overlappende segmenter af proberne E4.3 og H1.6;

figur 5 er en RNA blot hybridisering analyse, der anvender genomiske og cDNA prober. Hybridisering til fibroblast, trachea (normal og CF), pancreas, lever, HL60, T84 og hjerne RNA er vist;

5 figur 6 er methyleringsstatus for den E4.3 klonede region ved 5' enden af CF genet;

figur 7 er et restriktionsmap af den CFTR cDNA, der viser opstilling af cDNA'en til de genomiske DNA fragmenter;

10

figur 8 er en RNA gel blot analyse, der viser hybridisering af en portion af CFTR cDNA'en (klon 10-1) til en 6,5 kb mRNA transcript i forskellige humane væv;

15

figur 9 er en DNA blot hybridiseringsanalyse, der viser hybridisering af CFTR cDNA klonerne til genomisk DNA spaltet med EcoRI og Hind III;

figur 10 er et primer ekstension eksperiment, der karakteriserer 5' og 3' enderne af CFTR cDNA'en;

20

figur 11 er en hydropati-profil og viser forudsagte sekundære strukturer af CFTR;

figur 12 er en dot matrix analyse af interne homologier i det forudsagte CFTR polypeptid;

25

figur 13 er en skematisk model af det forudsagte CFTR protein;

figur 14 er et skematisk diagram af restriktionsfragment-længde polymorfismerne (RFLP'er) tæt forbundet til CF genet, hvor den omvendte trekant viser lokaliseringen af F508 3 basepar deletionen;

30

figur 15 repræsenterer påvisningen af F508 mutationen ved hjælp af oligonucleotid-hybridisering med Probe N påvisning af den normale sekvens og Probe F påvisning af den CF mutante sekvens;

35

figur 16 repræsenterer opstilling af de mest konserverede segmenter af de ekstenderede NBF'er af CFTR med sammenlignelige regioner af andre proteiner;

5 figur 17 er DNA sekvensen omkring F508 deletionen;

figur 18 er en repræsentation af den nucleotid sekventering gel, der viser DNA sekvensen ved F508 deletionen;

10 figur 19a og 19b er Coomassie Blå farvede polyacrylamidgeler efter elektroforese af protein fra bakterielle lysater (JM 101), hvilke bakterier blev transformeret med pGEX plasmiderne;

15 figur 20 er immunoblots af bakterielle lysater, der indeholder fusionsprotein #1 (i tabel 8) med præ-immun-sera og immun-sera fra to forskellige kaniner.

figur 21 er et immunoblot af T-84 membraner under anvendelse af immun-serum fra kanin #1 i figur 20; og

20 figur 22 er immunodot blots probet med præ-immun-sera og immun-sera fra en kanin immuniseret med KLH konjugatet af peptid #2 fra tabel 8.

Med henblik på at den foreliggende opfindelse kan forstås mere fuldstændigt, vil udførelsesformer deraf nu blive beskrevet blot som illustration.

25

## 1. DEFINITIONER

30 Med henblik på at fremme oversigt over de forskellige udførelsesformer ifølge opfindelsen og en forståelse af forskellige elementer og bestanddele anvendt til at udføre den foreliggende opfindelse og anvendelse deraf, er den følgende definition af udtryk anvendt i den foreliggende opfindelse beskrivelse som følger:

35

CF - cystisk fibrose

CF bærer - en tilsyneladende rask person, hvis kromosomer indeholder et mutant CF gen, der kan overføres til afkommet af den pågældende person.



CF patient - en person, som bærer et mutant CF gen på hvert kromosom, således at de udviser de kliniske symptomer på cystisk fibrose.

5 CF gen - det gen, hvis mutante former er associeret med sygdommen cystisk fibrose. Denne definition skal forstås at indbefatte de forskellige sekvenspolymorfismer, der eksisterer, hvor nucleotidsubstitutioner i gensekvensen ikke påvirker den essentielle funktion af genproduktet. Dette udtryk angår primært en isoleret kodende sekvens, men kan også indbefatte nogle eller alle de flankerende regulatoriske elementer og/eller introns.

10

CF - PI - cystisk fibrose pancreatisk insufficient, den større kliniske subgruppe af cystisk fibrose patienter, kendetegnet ved insufficient pancreatisk eksokrin funktion.

15

CF - PS - cystisk fibrose pancreatisk sufficient, en klinisk subgruppe af cystisk fibrose patienter med tilstrækkelig pancreatisk eksokrin funktion for normal fordøjelse af næring.

20

CFTR - cystisk fibrose transmembran konduktans regulator protein, kodet af CF genet. Denne definition indbefatter proteinet som isoleret fra humane eller animalske kilder, som produceret ved hjælp af rekombinante organismer, og som kemisk eller enzymatisk syntetiseret. Denne definition skal forstås at indbefatte de forskellige polymorfe former af det protein, hvor aminosyresubstitutioner i de variable regioner af sekvensen ikke påvirker den essentielle funktion af proteinet eller dets hydropatiske profil eller sekundære eller tertiære struktur.

25

DNA - standard nomenklatur anvendes til at identificere baserne.

30

Intronless DNA – et stykke af DNA, der mangler interne non-kodende segmenter, for eksempel cDNA.

IRP locus sekvens - (protoonkogen int-1 relateret), et gen lokaliseret nær CF genet.

35

Mutant CFTR - et protein, der er særdeles analogt med CFTR for så vidt angår primær, sekundær og tertiær struktur, men hvor et lille antal af

aminosyresubstitutioner og/eller deletioner og/eller insertioner resulterer i forringelse af dets essentielle funktion, således at organismer, hvis epiteliale celler udtrykker mutant CFTR frem for CFTR, udviser symptomerne på cystisk fibrose.

5

mCF - et muse gen ortologt med de humane CF gen

NBF'er - nucleotid (ATP) bindingsfolder

10

ORF - åben læseramme

PCR - polymerasekæde reaktion

15

Protein - standard enkelt bogstav nomenklatur anvendes til at identificere aminosyrerne

R-domæne - et stærkt ladet cytoplasmatisk domæne fra CFTR proteinet

RSV - Rous Sarcom Virus

20

SAP - surfactant protein

RFLP - restriktionsfragment længde polymorfisme

25

## 2. ISOLERING AF CF GENET

30

Under anvendelse af kromosom walking, jumping og cDNA hybridisering er DNA sekvenser, der omfatter > 500 kilobasepar (kb,) blevet isoleret fra en region på den lange arm af humant kromosom 7, der indeholder cystisk fibrose (CF) genet. Adskillige transkriberede sekvenser og konserverede segmenter er blevet identificeret i denne region. Én af disse svarer til CF genet og udspænder ca. 250 kb af genomisk DNA. Overlappende komplementære DNA (cDNA) kloner er blevet isoleret fra epiteliale celle biblioteker med et genomisk DNA segment, der indeholder en portion af cystisk fibrose genet. Nucleotid-sekvensen af den isolerede cDNA er vist i figur 1. I hver række af de respektive sekvenser er den nedre række en liste over standard nomenklatur for nucleotid-sekvensen. Den øvre række i hver respektive række af sekvenser

35

er standard enkelt bogstav nomenklatur for aminosyren svarende til den respektive codon.

5 Transcripter af ca. 6.500 nucleotider i størrelse er påviselige i berørte væv i patienter med CF. Baseret på den isolerede nucleotid-sekvens består det forudsagte protein af to lignende regioner, der hver indeholder et første domæne, der har egenskaber, som er konsistente med membranassociation, og et andet domæne, som er formodet at være involveret i ATP binding.

10 En 3 bp deletion, der resulterer i udeladelsen af en phenylalanin rest ved centeret af det første forudsagte nucleotid-bindende domæne (aminosyreposition 508 af CF gen produktet) er blevet påvist i CF patienter. Denne mutation i den normal DNA sekvens i figur 1 svarer til ca. 70 % af mutationerne i cystisk fibrose patienter. Ekstenderet haplotype-data baseret  
15 på DNA markører tæt forbundet til det formodede sygdoms gen tyder på, at det resterende af CF mutant gen puljen består af multiple, forskellige mutationer. Et lille sæt af disse sidstnævnte mutant alleler (ca. 8 %) kan bibringe residual pancreatisk eksokrin funktion i en subgruppe af patienter, som er pancreatisk sufficente.

20

### 2.1 KROMOSOM WALKING OG JUMPING

Store mængder af den DNA, der omgiver de D7S122 og D75340 bindende regioner af Rommens et al supra blev gennem søgt for kandidat gen  
25 sekvenser. Derudover blev traditionelle kromosom walking fremgangsmåder, kromosom jumping teknikker benyttet til at accelerere søgningsmetoden. Fra hvert jump endepunkt kunne en ny bidirektionel walk initieres. Sekventielle walks standset af "ikke-klonbare" regioner, ofte omfattet i det mammale genom, kunne omgås ved hjælp af kromosom jumping.

30

Det anvendte kromosom jumping bibliotek er blevet beskrevet tidligere [Collins et al, Science 235, 1046 (1987); Ianuzzi et al, Am. J. Hum. Genet. 44, 695 (1989)]. Det oprindelige bibliotek blev fremstillet fra en præparativ pulset felt gel, og var tiltænkt at indeholde partielle EcoR1 fragmenter på 70 -130 kb; efterfølgende erfaring med dette bibliotek viser, at mindre fragmenter også var  
35 repræsenteret, og jump størrelser på 25 - 110 kb er blevet fundet. Biblioteket blev udpladet på sup vært MC1061 og screenet ved hjælp af standard

teknikker, [Maniatis et al]. Positive kloner blev subklonet ind i pBR $\Delta$ 23Ava og begyndelsen og enden af jump'et identificeret ved hjælp af EcoR1 og Ava 1 spaltning, som beskrevet i Collins, Genome analysis: A practical approach (IRL, London, 1988), pp. 73-94). For hver klon blev et fragment fra enden af  
5 jump'et blev kontrolleret for at bekræfte dets lokalisering på kromosom 7. Den tilgrænsende kromosom region dækket af kromosom walking og jumping var omtrentligt 250 kb. Retning for jumps'ene blev biased ved omhyggeligt valg af prober, som beskrevet af Collins et al og Ianuzzi et al, supra. Hele den klonede region, herunder sekvenserne isoleret med anvendelsen af CF gen cDNA'en,  
10 er ca. 500 kb.

Den skematiske repræsentation af kromosom walking og jumping strategien er illustreret i figur 2. CF gen exoner er markeret ved hjælp af romertal i denne figur. Horisontale linjer over map'et viser walk-trin, mens buerne over map'et  
15 viser jump-trin. Figuren skrider frem fra venstre til højre i hver af seks rækker med retningen af ender hen imod 7cen og 7qter som vist. Restriktionsmap'et for enzymerne EcoRI, HindIII og BamHI er vist oven for den fuldt optrukne linje, der udspænder hele den klonede region. Restriktionssites vist med pile frem for lodrette linjer viser sites, der ikke er blevet utvetydigt positioneret.  
20 Yderligere restriktionssites for andre enzymer er vist nedenfor linjen. Gaps i den klonede region er markeret ved hjælp af ||. Disse forekommer kun i den portion, der er påvist af cDNA kloner af CF transcript'en. Disse gaps er sandsynligvis ikke store baseret på pulset felt mapping af regionen. Walking klonerne, som markeret ved hjælp af horisontale pile over map'et, har retningen af den pil, der viser walking fremskridtet opnået med hver klon.  
25 Cosmid kloner begynder med bogstavet c; alle andre kloner er fag. Cosmid CF26 viste sig at være en kimære; den stiplede del er afledt fra et anderledes genomisk fragment på et andet kromosom. Romertal I til og med XXIV viser lokaliseringen af exoner fra CF genet. De horisontale boxe vist oven for linjen er prober anvendt under eksperimenterne. Tre af proberne repræsenterer uafhængig subkloning af fragmenter tidligere identificeret til at påvise polymorfismer i denne region: H2.3A svarer til probe XV2C [X. Estivill et al, Nature, 326: 840 (1987)], probe E1 svarer til KM19 (Estivill, supra), og probe E4.1 svarer til Mp6d.9 [X. Estivill et al. Am. J. Hum. Genet. 44, 704 (1989)]. G-  
30 2 er et subfragment af E6, der påviser en transkriberet sekvens. R161, R159, og R160 er syntetiske oligonucleotider konstrueret fra dele af det sekventerede IRP locus sekventeret [B. J. Wainwright et al, EMBO J., 7:

1743(1988)], hvilket viser lokaliseringen af denne transcript på det genomiske map.

5 Eftersom de to uafhængigt isolerede DNA markører, D7S122 (pH131) og D7S340 (TM58), kun var ca. 10 kb fra hinanden (figur 2), blev walks'ene og jumps'ene i det væsentlige initieret fra et enkelt punkt. Retningen for walking og jumping med hensyn til MET og D7S8 blev dernæst etableret med krydsningen af adskillige sjældent-spaltende restriktionsendonuclease genkendelsessites (såsom de for Xho I, Nru I og Not I, se figur 2) og med  
10 henvisning til long range område fysisk map af J. M. Rommens et al. Am. J. Hum. Genet., i trykken; A. M. Poustka, et al, Genomics 2, 337 (1988); M. L. Drumm et al. Genomics 2, 346 (1988). Data fra pulset felt mapping afslørede også, at Not I site't identificeret af de foreliggende opfindere af den foreliggende opfindelse (se figur 2, position 113 kb) svarede til den, der  
15 tidligere var fundet associeret IRP locus'et (Estivill et al 1987, supra). Eftersom efterfølgende genetiske studier viste, at CF mest sandsynligt var lokaliseret mellem IRP og D7S8 [M. Farrall et al, Am. J. Hum. Genet. 43, 471 (1988), B.-S. Kerem et al. Am. J. Hum. Genet. 44, 827 (1989)], blev walking og jumping bestræbelserne fortsat udelukkende hen imod kloning af dette interval. Det erkendes imidlertid, at andre kodende regioner, som identificeret i figur 2, for eksempel G-2, CF14 og CF16, var lokaliseret og undersøgt intensivt. Sådanne omfattende undersøgelser af disse øvrige regioner afslørede, at de ikke var CF genet baseret på genetiske data og sekvensanalyse. Givet manglen på viden om lokaliseringen af CF genet og dets karakteristika, fremmede den  
20 omfattende og tidskrævende undersøgelse af de nærvæd formodede kodende regioner ikke retningen for søgning efter CF genet. Disse undersøgelser var imidlertid nødvendige med henblik på at udelukke muligheden for, at CF genet er i de pågældende regioner.

30 Tre regioner i 280 kb segmentet viste sig ikke at kunne udvindes let i de amplificerede genomiske biblioteker, der initialt blev anvendt. Disse mindre klonbare regioner var lokaliseret nær DNA segmenterne H2.3A og X.6, og lige hinsides cosmid cW44, ved position henholdsvis 75 - 100 kb, 205 - 225 kb, og 275 - 285 kb i figur 2. De rekombinante kloner nær H2.3A viste sig at være meget ustabile med dramatiske omflytninger efter blot nogle få passager af bakteriel dyrkning. Med henblik på at udfylde de dannede gaps blev primære walking biblioteker konstrueret under anvendelse af specielle vært-vektor-

35

systemer, som er blevet rapporteret at tillade opformering af ustabile sekvenser (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)). Selv om regionen nær cosmid cW44 stadig ikke er udvundet, blev regionen nær X.6 succesfuldt reddet med disse biblioteker.

## 2.2 KONSTRUKTION AF GENOMISKE BIBLIOTEKER

10 Genomiske biblioteker blev konstrueret efter procedurer beskrevet i Maniatis, et al, Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 1982) og er listet i tabel 1. Dette indbefatter otte fag biblioteker, hvoraf ét blev tilvejebragt af T. Maniatis [Fritsch et al, Cell, 19: 959 (1980)]; resten blev konstrueret som en del af det foreliggende ifølge procedurer beskrevet i Maniatis et al, supra. Fire fag biblioteker blev klonet i 15  $\lambda$ DASH (kommercielt tilgængelige fra Stratagene) og tre i  $\lambda$ FFIX (kommercielt tilgængelige fra Stratagene), med vektor-arme tilvejebragt af producenten. Ét  $\lambda$ DASH bibliotek blev konstrueret fra Sau 3A-partielt spaltet DNA fra en human-hamster hybrid, der indeholdt humant kromosom 7 (4AF/102/K015) 20 [Rommens et al Am. J. Hum. Genet 43, 4 (1988)], og andre biblioteker fra partiel Sau3A, total BamHI eller total EcoRI spaltning af humant perifert blod eller lymfoblastoid DNA. For at undgå tab af ustabile sekvenser blev fem af fag bibliotekerne opformeret på de rekombinations-deficiente værter DB1316 (recD<sup>-</sup>), CES 200 (recBC<sup>-</sup>) (Wyman et al, supra, Wertman et al supra, Wyman et al supra); eller TAP90 [Patterson et al Nucleic Acids Res. 15:6298 (1987)]. 25 Tre cosmid biblioteker blev dernæst konstrueret. I ét blev vektoren pCV108 (Lau et al Proc. Natl. Acad. Sci USA 80:5225 (1983)) anvendt til at kloner partielt spaltet (Sau 3A) DNA fra 4AF/102/K015 [Rommens et al Am. J. Hum. Genet. 43:4 (1988)]. Et andet cosmid bibliotek blev fremstillet ved kloning af 30 partielt spaltet (Mbo I) human lymfoblastoid DNA ind i vektoren pWE-IL2R, fremstillet ved insertion af RSV (Rous Sarcoma Virus) promotor-drevet cDNA'en for interleukin-2 receptor kæden (leveret af M. Fordis og B. Howard) i stedet for neo-resistens genet af pWE15t [Wahl et al Proc. Natl. Acad. Sci. USA 84:2160 (1987)]. Et yderligere partielt Mbo I cosmid bibliotek blev 35 fremstillet i vektoren pWE-IL2-Sal, frembragt ved insertion en Sal I linker i Bam HI kloning site't af pWE-EL2R (M. Drums, upublicerede data); dette tillader anvendelsen af den partielle udfyldningsteknik til ligering af Sal I og Mbo I

ender, hvilket forhindrer tandem insertioner (Zabarovsky, et al Gene 42:19 (1986)]. Cosmid biblioteker blev opformeret i E. coli vært stammer DH1 eller 490A [M. Steinmetz, A. Winoto, K. Minard, L. Hood, Cell 28, 489 (1982)].

5

TABEL 1

## GENOMISKE BIBLIOTEKER

Vektor	Kilde til human DNA	Vært	Kompleksitet	Ref
$\lambda$ Charon 4A	HaeIII/AluI-partielt spaltet total human lever DNA	LE392	$1 \times 10^6$ (amplificeret)	Lawn et al 1980
PCV108	Sau3a-partielt spaltet DNA fra 4AF/KO15	DK1	$3 \times 10^6$ (amplificeret)	
$\lambda$ dash	Sau3A-partielt spaltet DNA fra 4AF/KO15	LE392	$1 \times 10^6$ (amplificeret)	
$\lambda$ dash	Sau3A-partielt spaltet total human perifert blod DNA	DB1316	$1,5 \times 10^6$	
$\lambda$ dash	BamHI-spaltet total human perifert blod DNA	DB1316	$1,5 \times 10^6$	
$\lambda$ dash	EcoRI-partielt spaltet total human perifert blod DNA	DB1316	$8 \times 10^6$	
$\lambda$ FIX	MboI-partielt spaltet human lymfoblastoid DNA	LE392	$1,5 \times 10^6$	
$\lambda$ FIX	MboI-partielt spaltet human lymfoblastoid DNA	CE200	$1,2 \times 10^6$	
$\lambda$ FIX	MboI-partielt spaltet human lymfoblastoid DNA	TAP90	$1,3 \times 10^6$	
pWE-IL2R	MboI-partielt spaltet human lymfoblastoid DNA	490A	$5 \times 10^5$	
pWE-IL2r-Sal	Mobol-partielt spaltet human lymfoblastoid DNA	<u>490A</u>	$1,2 \times 10^6$	
$\lambda$ Ch3A $\Delta$ lac (jumping)	EcoRI-partielt spaltet (24-110 kb) human lymfoblastoid DNA	MC1061	$3 \times 10^6$	Collins Supra and Iannuzzi et al supra

Tre af fag bibliotekerne blev opformeret og amplificeret i *E. coli* bakteriel stamme LE392. Fire efterfølgende biblioteker blev udpladet på de rekombinationsdeficiente værter DB1316 (*recD*<sup>-</sup>) eller CES200 (*recBC*<sup>-</sup>) [Wyman 1985, supra; Wertman 1986, supra; og Wyman 1986, supra eller i ét tilfælde TAP90 [T.A. Patterson and M. Dean, Nucleic Acids Research 15, 6298 (1987)].

Enkelt kopi DNA segmenter (fri af repetitive elementer) nær enderne af hver fag eller cosmid insertion blev oprenset og anvendt som prober til biblioteksscreening med henblik på isolere overlappende DNA fragmenter ved hjælp af standard procedurer. (Maniatis, et al, supra).

1-2 x 10<sup>6</sup> fag kloner blev udpladet på 25-30 150 nm petriskåle med den passende indikator bakterielle vært og inkuberet ved 37 °C i 10 - 16 timer. Dobbeltbestemmelse "lifts" blev fremstillet for hver plade med nitrocellulose- eller nylon-membraner, præhybridiseret og hybridiseret under beskrevne betingelser (Rommens et al, 1988, supra). Prober blev mærket med <sup>32</sup>P til en specifik aktivitet på >5 x 10<sup>8</sup> cpm/µg under anvendelse af random priming proceduren [A. P. Feinberg og B. Vogelstein, Anal. Biochem. 132, 6 (1983)]. Cosmid biblioteket blev spredt ud på ampicillin-holdige plader og screenet på en lignende måde.

DNA prober, der gav høje baggrundssignaler, kunne ofte anvendes mere succesfuldt ved præ-annealing af den kogte probe med 250 µg/ml sheared denatureret placental DNA i 60 minutter forud for tilsætning af proben til hybridiseringssækken.

For hvert walk step blev identiteten af det klonede DNA fragment bestemt ved hjælp af hybridisering med et somatisk celle hybrid panel med henblik på at bekræfte dets kromosomale lokalisering og ved hjælp af restriktionsmapping og Southern blot analyse for at bekræfte dets kolinearitet med genomet.

Den totale kombinerede klonede region af de isolerede genomiske DNA sekvenser isoleret og de overlappende cDNA kloner, ekstenderede > 500 kb. For at sikre, at DNA segmenterne isoleret ved hjælp af kromosom walking og jumping procedurerne var ko-lineær med den genomiske sekvens, blev hvert segment undersøgt ved hjælp af:



(a) hybridiseringsanalyse med human-gnaver somatisk hybride cellelinjer med henblik på at bekræfte kromosom 7 lokalisering,

(b) pulset felt gel elektroforese, og

5

(c) sammenligning af restriktionsmap'et af den klonede DNA til den for af den genomiske DNA.

Således blev en enkelt kopi humane DNA sekvenser isoleret fra hver rekombinant fag og cosmid klon og anvendt som prober i hver af disse hybridiseringsanalyser som udført ved hjælp af proceduren af Maniatis, et al supra.

10

Mens størstedelen af fag og cosmid isolater repræsenterede korrekte walk og jump kloner, var nogle få resultat af kloningsartefacter eller kryds-hybridiserende sekvenser fra andre regioner i det humane genom, eller fra hamster genomet i tilfælde, hvor bibliotekerne blev afledt fra en human-hamster hybrid cellelinje. Bekræftelse af korrekt lokalisering var særligt vigtig for kloner isoleret ved hjælp af kromosom jumping. Mange jump kloner blev overvejet og resulterede i non-konklusiv information, der førte til retningen for udforskning fra genet.

15

20

### 2.3 BEKRÆFTELSE AF RESTRIKTIONSMAP'ET

Yderligere bekræftelse af det totale fysiske map af de overlappende kloner blev opnået ved long range restriktionsmapping analyse med anvendelsen af pulset felt gel elektroforese (J. M. Rommens, et al. Am. J. Hum. Genet., i trykken, A. M. Poustka et al, 1988, supra M. L. Drumm et al, 1988 supra).

25

Figur 3A til 3E viser fundene fra long range restriktionsmapping studiet, hvor en skematisk repræsentation af regionen er givet i Panel E. DNA fra human-hamster cellelinjen 4AF/102/K015 blev spaltet med enzymerne (A) Sal I, (B) Xho I, (C) Sfi I og (D) Nae I, adskilt ved hjælp af pulset felt gel elektroforese og overført til Zetaprobe™ (BioRad). For hvert enzym blev et enkelt blot sekventielt hybridiseret med proberne vist under hver af panelerne i figur A til D, med stripping af blot'et mellem hybridiseringer. Symbolerne for hvert enzym i figur 3E er: A, Nae I; B, Bss HII; F, Sfi I; L, Sal I; M, Mlu I; N, Not I; R, Nru I;

30

35

og X, Xho 1. C svarer til kompressionszone-regionen af gelen. DNA præparationer, restriktionsspaltning og krydset felt gel elektroforese fremgangsmåder er blevet beskrevet (Rommens et al, i trykken, supra). Gelerne i figur 3 blev kørt i 0,5 X TBE ved 7 volt/cm i 20 timer med switching lineært opkørt fra 10 - 40 sekunder for (A), (B) og (C), og ved 8 volt/cm i 20 timer med switching opkørt lineært fra 50 - 150 sekunder for (D). Skematiske fortolkninger af hybridiseringsmønsteret er givet under hvert panel. Fragmentlængder er i kilobaser og blev størrelsesbestemt ved sammenligning med oligomeriseret bakteriofag  $\lambda$ DNA og Saccharomyces cerevisiae kromosomer.

H4.0, J44, EG1.4 er genomiske prober frembragt fra walking og jumping eksperimenterne (se figur 2). J30 er blevet isoleret ved hjælp af fire konsekutive jumps fra D7S8 (Collins et al, 1987, supra; Ianuzzi et al, 1989, supra; M. Dean, et al, indleveret til publikation). 10-1, B.75 og CE1.5/1.0 er cDNA prober, som dækker forskellige regioner af CF transcriptet: 10-1 indeholder exonerne I - VI, B.75 indeholder exonerne V - XII, og CE1.5/1.0 indeholder exonerne XII - XXIV. Vist i figur 3E er et sammensat map af hele MET - D7S8 intervallet. Den indrammede region viser segmentet klonet ved hjælp af walking og jumping, og den stiplede portion viser regionen dækket af CF transcript'en. Den CpG-rige region associeret med D7S23 locus'et (Estivill et al, 1987, supra) er ved Not I site't vist i parentes. Dette og andre sites vist i parenteser eller firkantede parenteser spalter ikke i 4AF/102/K015, men er blevet observeret i humane lymfoblast cellelinjer.

#### 2.4 IDENTIFICERING AF CF GEN

Baseret på fundene ved long range restriktionsmapping detaljeret beskrevet ovenfor blev det bestemt, at hele CF genet er indeholdt på et 380 kb Sal I fragment. Opstilling af restriktionssites'ene afledt fra pulset felt gel analyse til de, der er identificeret i de partielt overlappende genomiske DNA kloner, afslørede, at størrelsen af CF genet var ca. 250 kb.

Det mest informative restriktionsenzym, der fungerede til at opstille map'et af de klonede DNA fragmenter og long range restriktionsmap'et, var Xho I; alle 9 Xho 1 sites'ene identificeret med de rekombinante DNA kloner så ud til at være modtagelige for mindst partiel spaltning i genomisk DNA (sammenlign

maps i figur 1 og 2). Derudover identificerede hybridiseringsanalyse med prober afledt fra 3' enden af CF genet 2 Sfil sites og bekræftede positionen for et forventet Nae I site.

5 Disse fund understøttede yderligere den konklusion, at DNA segmenterne isoleret ved hjælp af kromosom walking og jumping procedurene var ko-lineær med den ægte sekvens.

#### 10 2.5 KRITERIER FOR IDENTIFICERING

Et positivt resultat baseret på ét eller flere af de følgende kriterier tydede på, at et klonet DNA segment kan indeholde kandidat gen sekvenser:

15 (a) påvisning af kryds-hybridiserende sekvenser i andre arter (eftersom mange gener udviser evolutionsmæssig konservering),

(b) identifikation af CpG øer, som ofte markerer 5' enden af vertebrate gener [A. P. Bird, Nature, 321, 209 (1986); M. Gardiner-Garden og M. Frommer, J. Mol. Biol. 196, 261 (1987)],

20 (c) undersøgelse af mulige mRNA transcripter i berørte væv hos CF patienter,

(d) isolation af korresponderende cDNA sekvenser,

25 (e) identifikation af åbne læserammer ved hjælp af direkte sekventering af klonede DNA segmenter.

30 Kryds-art hybridisering viste stærk sekvenskonservering mellem human og bovin DNA, når CF14, E4.3 og H1.6 blev anvendt som prober, hvilke resultater er vist i figur 4A, 4B og 4C.

Humane, bovine, muse, hamster og kylling genomiske DNA'er blev spaltet med Eco RI (R), Hind III (H) og Pst I (P), underkastet elektroforese og blottet til Zetabind™ (BioRad). Hybridiseringsprocedurene af Rommens et al, 1988, supra, blev anvendt med den mest stringente vask ved 55 °C, 0,2 X SSC og 0,1 % SDS. Proberne anvendt til hybridisering, i figur 4, indbefattede: (A) helt

cosmid CF14, (B) E4.3, (C) H1.6. I det skematiske i figur (D), viser den skyggebelagte region området for kryds-art konservering.

5 Det faktum, at forskellige undergrupper af bånd blev påvist i bovin DNA med disse to overlappende DNA segmenter (H1.6 og E4.3), tydede på, at de konserverede sekvenser var lokaliseret ved grænserne for den overlappede region (figur 4(D)). Når disse DNA segmenter blev anvendt til at påvise RNA transcripter fra en række væv, blev intet hybridiseringssignal påvist. I et forsøg på at forstå den krydshybridiserende region og med henblik på at identificere mulige åbne læserammer blev DNA sekvenserne af hele H1.6 og en del af E4.3 fragmentet bestemt. Resultaterne viste, at, bortset fra en lang strækning af CG-rig sekvens, der indeholdt genkendelsessites'ene for to restriktionsenzymmer (Bss HII og Sac II), ofte fundet associeret med undermethylerede CpG øer, var der kun korte åbne læserammer, hvilket ikke 10 let kunne forklare de stærke kryds-art hybridiseringssignaler.

For at undersøge methyleringsstatus for denne særdeles CpG-rige region afsløret ved hjælp af sekventering blev genomisk DNA prøver fremstillet fra fibroblaster og lymfoblaster blev spaltet med restriktionsenzymmerne Hpa II og Msp I og analyseret ved hjælp af gel blot hybridisering. Enzymet Hpa II spalter kun DNA sekvensen 5'-CCGG-3', når den anden cytosin er umethyleret, mens Msp I spalter denne sekvens uanset tilstanden for methylering. Små DNA fragmenter blev frembragt ved hjælp af begge enzymer, hvilket viser, at denne CpG-rige region virkelig er undermethyleret i genomisk DNA. Gel-blot hybridiseringen med E4.3 segmentet (figur 6) afslører meget små hybridiseringsfragmenter med begge enzymer, hvilket viser tilstedeværelsen af en hypomethyleret CpG ø. 20

De ovennævnte resultater tyder stærkt på tilstedeværelsen af en kodende region ved dette locus. To DNA segmenter (E4.3 og H1.6), som påviste kryds-art hybridiseringssignaler fra dette område blev anvendt som prober til at screene cDNA biblioteker frembragt fra adskillige væv og celletyper. 30

cDNA biblioteker fra dyrkede epiteliale celler blev fremstillet som følger. Svedkirtel-celler afledt fra et non-CF individ og fra en CF patient blev dyrket til første passage som beskrevet [G. Collie et al, *In Vitro Cell. Dev. Biol.* 21, 592, 1985). Tilstedeværelsen af udadtil afhjælpnde kanaler blev bekræftet i disse 35

celler (J.A. Tabcharani, T.J. Jensen, J. R. Riordan, J. W. Hanrahan, J. Memb. Biol., i trykken), men CF cellerne var ufølsomme for aktivering af cyklisk AMP (T.J. Jensen, J.W. Hanrahan, J. A. Tabcharani, M. Buchwald and J. R. Riordan, Pediatric Pulmonology, Supplement 2, 100, 1988). RNA blev isoleret  
5 fra dem ved hjælp af fremgangsmåden af J. M. Chirgwin et al (Biochemistry 18, 5294, 1979). Poly A+RNA blev selekteret (H. Aviv og P. Leder, Proc. Natl. Acad. Sci. USA 69, 1408, 1972) og anvendt som template for syntesen af cDNA med oligo (dT) 12-18 som en primer. Den anden streng blev syntetiseret ifølge Gubler og Hoffman (Gene 25, 263, 1983). Denne blev methyleret med  
10 Eco RI methylase og ender blev gjort flush med T4 DNA polymerase. Phosphorylerede Eco RI linkere blev ligeret til cDNA'en og restrikted med Eco RI. Fjernelse af overskydende linkere og partiel størrelsesfraktionering blev opnået ved hjælp af Biogel A-50 kromatografi. CDNA'erne blev dernæst ligeret til Eco RI site't af den kommercielt tilgængelige lambda ZAP. Rekombinanter  
15 blev pakket og opformeret i E. coli BB4. Portioner af pakningsblandingerne blev amplificeret og det resterende tilbageholdt for screening forud for amplificering. De samme procedurer blev anvendt til at konstruere et bibliotek fra RNA isoleret fra præ-konfluerende kulturer af T-84 coloncarcinom cellelinjen (Dharmasathaphorn, K. et al. Am. J. Physiol. 246, G204, 1984).  
20 Antallene af uafhængige rekombinanter i de tre biblioteker var:  $2 \times 10^6$  for non-CF svedkirtel-cellerne,  $4,5 \times 10^6$  for CF svedkirtel-cellerne og  $3,2 \times 10^6$  fra T-84 celler. Disse fager blev udpladet ved 50.000 per 15 cm plade og plaque-lifts frembragt under anvendelse af nylon-membraner (Biodyne) og probet med DNA fragmenter mærket med  $^{32}\text{P}$  under anvendelse af DNA polymerase I og  
25 en random blanding af oligonucleotider som primer. Hybridiseringsbetingelser var ifølge G.M. Wahl og S.L. Berger (Meth. Enzymol. 152, 415, 1987). Bluescript™ plasmider blev reddet fra plaque-oprensede kloner ved udsækning med M13 hjælper fag. Lunge og pancreas bibliotekerne blev fremskaffet fra Clontech Lab Inc. med rapporterede størrelser på  $1,4 \times 10^6$  og  $1,7 \times 10^6$   
30 uafhængige kloner.

Efter screening af 7 forskellige biblioteker, der hver indeholdt  $1 \times 10^5$  -  $5 \times 10^6$  uafhængige kloner, 1 enkelt klon (identificeret som 10-1) blev isoleret med H1.6 fra et cDNA bibliotek frembragt fra de dyrkede svedkirtel-epitelceller fra et  
35 uberørt (non-CF) individ.

DNA sekventeringsanalyse viste, at 10-1 indeholdt en insertion af 920 bp i størrelse og én potentiel, lang åben læseramme (ORF). Eftersom den ende af sekvensen delte perfekt sekvensidentitet med H1.6, blev det konkluderet, at cDNA klonen sandsynligvis var afledt fra denne region. Den fælles DNA sekvens var imidlertid kun 113 bp lang (se figur 1 og 7). Som detaljeret beskrevet nedenfor svarede denne sekvens i realiteten til den 5'-nærmeste exon af det formodede CF gen. Den korte sekvens overlapper således forklaret de svage hybridiseringssignaler i bibliotekscreening og manglende evne til at påvise transcripter i RNA gel-blot analyse. Derudover blev orienteringen af transcriptions-unit'en tentativt etableret på basis af opstilling af den genomiske DNA sekvens med den formodede ORF af 10-1.

Eftersom den korresponderende transcript blev estimeret til at være ca. 6500 nucleotider i længde ved hjælp af RNA gel-blot hybridiseringseksperimenter, var yderligere cDNA bibliotek screening påkrævet med henblik på at klonere det resterende af den kodende region. Som en følge af adskillige successive screeninger med cDNA biblioteker frembragt fra coloncarcinom cellelinjen T84, normale og CF svedkirtel-celler, pancreas og voksne lunger, blev der isoleret 18 yderligere kloner (figur 7, som efterfølgende drøftet mere detaljeret). DNA sekvens analyse afslørede, at ingen af disse cDNA kloner svarede til længden af den observerede transcript, men det var muligt at aflede en konsensus sekvens baseret på overlappende regioner. Yderligere cDNA kloner svarende til 5' og 3' enderne af transcripten blev afledt fra 5' og 3' primer-ekstension eksperimenter. Tilsammen udspænder disse kloner en total på omtrentligt 6,1 kb og indeholder en ORF, der er i stand til at kode et polypeptid på 1480 aminosyrerester (figur 1).

Det var usædvanligt at observere, at de fleste af cDNA klonerne isoleret her indeholdt sekvens-insertioner ved forskellige lokaliseringer af restriktionsmap'et i figur 7. Map'et viser detaljer for den genomiske struktur af CF genet. Exon/intron grænser er givet, hvor alle isolerede cDNA kloner er skematisk repræsenteret på den øvre halvdel af figuren. Mange af disse ekstra sekvenser svarede klart til intron regioner omvendt transkriberet under konstruktionen af cDNA'en, som afsløret efter opstilling med genomiske DNA sekvenser.

Eftersom antallet af rekombinante cDNA kloner for CF genet påvist i bibliotek-screeningen var meget mindre, end hvad der ville være blevet forventet fra rigeligheden af transcript estimeret fra RNA hybridisering eksperimenter, forekom det sandsynligt, at de kloner, der indeholdt aberrante strukturer blev tilbageholdt med præference, mens de rigtige kloner blev tabt under opformering. Konsistent med denne fortolkning blev der observeret ringe vækst for størstedelen af de rekombinante kloner isoleret i dette studium, uanset den anvendte vektor.

Procedurerne anvendt til at opnå 5' og 3' enderne af cDNA'en var svarende til de, der er beskrevet (M. Frohman et al, Proc. Nat. Acad. Sci, USA, 85, 8998 - 9002, 1988). For 5' ende klonerne blev total pancreas og T84 poly A + RNA prøver omvendt transkriberet under anvendelse af en primer, (10b), der er specifik for exon 2 på tilsvarende måde, som det er blevet beskrevet for primer ekstension reaktionen, bortset fra at radioaktiv tracer blev indbefattet i reaktionen. Fraktionerne opsamlet fra en agarose bead søjle af den første streng syntese blev undersøgt ved hjælp af polymerase kæde reaktion (PCR) af eluerede fraktioner. De anvendte oligonucleotider var inden for 10-1 sekvensen (145 nucleotider fra hinanden) lige 5' for ekstension primeren. De tidligste fraktioner, der gav PCR produkt, blev puljet og koncentreret ved hjælp af inddampning og efterfølgende tallet med terminal deoxynucleotidyl transferase (BRL Labs.) og dATP som anbefalet af leverandøren (BRL Labs). En anden streng syntese blev dernæst udført med Taq Polymerase (Cetus, AmpliTaq™) under anvendelse af et oligonucleotid, der indeholdt en tallet linker sekvens 5'CGGAATTCTCGAGATC(T)<sub>12</sub>3'.

Amplificering af et forankret (PCR) eksperiment under anvendelse af linker sekvensen og en primer lige intern for den ekstensionsprimer, som besad Eco RI restriktionssite't ved dens 5' ende, blev dernæst udført. Efter restriktion med enzymerne Eco RI og Bgl II og agarose gel oprensning blev størrelsesselekerede produkter klonet ind i plasmidet Bluescript KS tilgængelig fra Stratagene ved hjælp af standard procedurer (Maniatis et al, supra). I det væsentlige indeholdt alle de udvundne kloner insertioner på mindre end 350 nucleotider. For at opnå 3' ende klonerne blev første streng cDNA fremstillet med omvendt transcription af 2 µg T84 poly A + RNA under anvendelse af det tallede linker oligonucleotid tidligere beskrevet med betingelser svarende til dem for primer ekstensionen. Amplificering ved hjælp

af PCR blev dernæst udført med linker oligonucleotidet og tre forskellige oligonucleotider svarende til kendte sekvenser af klon T16-4.5. En præparativ skala reaktion (2 x 100 ml) blev udført med ét af disse oligonucleotider med sekvensen 5'ATGAAGTcCAAGGATTTAG3'.

5

Dette oligonucleotid er ca. 70 nucleotider opstrøms for et Hind III site inden for den kendte sekvens af T16-4.5. Restriktion af PCR produktet med Hind III og Xho I blev efterfulgt af agarose gel oprensning for at størrelsesselekttere et bånd ved 1,0 - 1,4 kb. Dette produkt blev dernæst klonet ind i plasmidet Bluescript KS tilgængeligt fra Stratagene. Omtrentligt 20 % af de opnåede kloner hybridiserede til 3' ende portionen af T16-4.5. 10/10 af plasmider isoleret fra disse kloner havde identiske restriktionsmaps med insertionsstørrelser på omtrentligt 1,2 kb. Alle PCR reaktionerne blev udført i 30 cyklusser i buffer foreslået af en enzym-leverandør.

15

En ekstensionsprimer positioneret 157 nt fra 5'enden af 10-1 klon blev anvendt til at identificere startpunktet for den formodede CF transcript. Primeren blev ende-mærket med  $\gamma$ [<sup>32</sup>P]ATP ved 5000 Curie/mmol og T4 polynucleotid kinase og oprenset ved hjælp af slyngning søjle gel filtration. Den radioaktivt mærkede primer blev dernæst annealet med 4 - 5  $\mu$ g poly A + RNA fremstillet fra T-84 coloncarcinom celler i 2 X omvendt transcriptase buffer i 2 timer ved 60 °C. Efter fortynding og tilsætning af AMV omvendt transcriptase (Life Sciences, Inc.) skred inkubering ved 41 °C frem i 1 time. Prøven blev dernæst justeret til 0,4 M NaOH og 20 mM EDTA og til sidst neutraliseret med NH<sub>4</sub>OAc, pH 4,6, phenol-ekstraheret, ethanol-præcipiteret, genopløst i buffer med formamid og analyseret på en polyacrylamid sekventering gel. Detaljer for disse fremgangsmåder er blevet beskrevet (Meth. Enzymol. 152, 1987, Ed. S. L. Berger, A. R. Kimmel, Academic Press, N.Y.).

20

25

Resultater af primer ekstension eksperimentet under anvendelse af en ekstension oligonucleotid primer, der startede 157 nucleotider fra 5' enden af 10-1, er vist i Panel A i figur 10. Ende-mærket  $\phi$ X174 bakteriofag spaltet med Hae III (BRL Labs) anvendes som størrelsesmarkør. To større produkter er observeret ved 216 og 100 nucleotider. Sekvensen svarende til 100 nucleotider i 10-1 svarer til en meget GC rig sekvens (11/12), hvilket tyder på, at denne kunne være et omvendt transcriptase pause site. De 5' forankrede PCR resultater er vist i panel B i figur 10. Den 1,4 % agarose gel, der er vist til

35



venstre, blev blottet og overført til Zeta probe X membran (Bio-Rad Lab). DNA gel blot hybridisering med radioaktivt mærket 10-1 er vist til højre. 5' ekstension produkterne ses at variere i størrelse fra 170 - 280 nt med det større produkt ved omtrentligt 200 nucleotider. PCR kontrol banen viser et fragment på 145 nucleotider. Det blev opnået ved at anvende test oligomererne inden for 10-1 sekvensen. De viste størrelsesmarkører svarer til størrelser på 154, 220/210, 298, 344, 394 nucleotider ("lkb ladder" fremskaffet fra BRL Lab).

10 Det skematiske vist under Panel B i figur 10 beskriver proceduren til opnåelse af dobbeltstrenget cDNA anvendt for amplificeringen og kloningen til at frembringe klonerne PA3-5 og TB2-7 vist i figur 7. De forankrede PCR eksperimenter til at karakterisere 3'enden er vist i panel C. Som vist i det skematiske under figur 10C, blev tre primere, hvis relative position i forhold til  
15 hinanden var kendt, anvendt til amplificering med omvendt transkriberet T84 RNA som beskrevet. Disse produkter blev adskilt på en 1 % agarose gel og blottet på nylon membran som beskrevet ovenfor. DNA-blot hybridisering med 3' portionen af T16-4.5 klonen gav bånd med størrelser, der svarede til afstanden mellem den anvendte specifikke oligomer og 3'enden af transcript'en. Disse bånd i banerne 1, 2a og 3 er vist skematisk under Panel C  
20 i figur 10. Båndet i bane 3 er svagt, eftersom kun 60 nucleotider af dette segment overlapper med den anvendte probe. Også vist i det skematiske og som vist i banen 2b er produktet frembragt ved hjælp af restriktion af det forankrede PCR produkt med henblik på at fremme kloning til frembringelse af  
25 THZ-4 klonen vist i figur 7.

DNA-blot hybridisering analyse af genomisk DNA spaltet med EcoRI og HindIII enzymer probet med portioner af cDNA'er, der udspænder hele transcript'en tyder på, at genet indeholder mindst 24 exoner nummereret som romertal I til  
30 og med XXIV (se figur 9). Disse svarer til numrene 1 til og med 24 vist i figur 7. Størrelsen af hvert bånd er givet i kb.

I figur 7 viser åbne boxe omtrentlige positioner af de 24 exoner, som er blevet identificeret ved hjælp af isolationen af > 22 kloner fra screeningen af cDNA biblioteker og fra forankrede PCR eksperimenter designet til at klonere 5' og 3' enderne. Længderne i kb af de EcoRI genomiske fragmenter påvist af hver  
35 exon er også vist. De skraverede boxe i figur 7 viser tilstedeværelsen af intron

sekvenser, og de stiplede boxe viser andre sekvenser. Vist nederst til venstre af den lukkede box er den relative position af klonen H1.6 anvendt til at påvise den første cDNA klon 10-1 fra blandt  $10^6$  fag af det normale svedkirtel bibliotek. Som vist i figur 4(D) og 7 overlapper den genomiske klon H1.6  
5 partielt med et EcoRI fragment på 4,3 kb. Alle de viste cDNA klone vist blev hybridiseret til genomisk DNA og/eller var fint restriktionsmappet. Eksempler på de restriktionssites, der forekommer inden for cDNA'erne og i de korresponderende genomiske fragmenter er vist.

10 Med reference til figur 9 indbefatter hybridiseringsanalysen prober; i.e. cDNA kloner 10-1 for panel A, T16-1 (3' portion) for panel B, T16-4.5 (central portion) for panel C og T16-4.5 (3' ende portion) for panel D. I panel A i figur 9 påviser cDNA proben 10-1 de genomiske bånd for exonerne I til og med VI. 3' portionen af T16-1 frembragt ved hjælp af NruI restriktion påviser exonerne IV  
15 til og med XIII som vist i Panel B. Denne probe overlapper partielt med 10-1. Panelerne C og D viser henholdsvis genomiske bånd påvist ved central og 3' ende EcoRI fragmenterne af klonen T16-4.5. To EcoRI sites forekommer inden for cDNA sekvensen og split exonerne XIII og XIX. Som vist ved hjælp af exonerne i parenteser, to genomiske EcoRI bånd svarer til hver af disse  
20 exoner. Krydshybridisering til andre genomiske fragmenter blev observeret. Disse bånd, markeret ved hjælp af N, er ikke af kromosom 7 oprindelse, eftersom de ikke fremkom i human-hamster hybrider, der indeholdt humant kromosom 7. Det svage bånd i panel D markeret ved hjælp af XI i parenteser formodes at være forårsaget af kryds-hybridiseringen af sekvenser på grund af  
25 intern homologi med cDNA'en.

Eftersom 10-1 påviste et stærkt bånd på gel blot hybridisering af RNA fra T-84 coloncarcinom cellelinjen, blev denne cDNA anvendt til at screene biblioteket konstrueret fra den pågældende kilde. Femten positive blev opnået, hvorfra  
30 klonerne T6, T6/20, T11, T16-1 og T13-1 blev oprenset og sekventeret. Rescreening af det samme bibliotek med et 0,75 kb Bam HI-Eco RI fragment fra 3' enden af T16-1 gav T16-4.5. Et 1,8 kb EcoRI fragment fra 3' enden af T16-4.5 gav T8-B3 og T12a, hvor sidstnævnte indeholdt et polyadenylationssignal og hale. Samtidigt blev et human lunge cDNA bibliotek screenet; mange kloner blev isoleret, herunder de, der er vist her med  
35 forstavelsen 'CDL'. Et pancreas bibliotek blev også screenet, hvilket gav klon CDPJ5.

Med henblik på at opnå kopier af denne transcript fra en CF patient blev et cDNA bibliotek fra RNA af svedkirtel-epitelceller fra en patient screenet med 0,75 kb Bam HI - Eco RI fragmentet fra 3' enden af T16-1, og klonerne C16-1 og C1-1/5, som dækkede alle undtagen exon 1, blev isoleret. Disse to kloner  
5 udviste begge en 3 bp deletion i exon 10, som ikke er til stede i nogen anden klon, der indeholdt den pågældende exon. Adskillige kloner, herunder CDLS26-1 fra lunge biblioteket og T6/20 og T13-1 isoleret fra T84, blev afledt fra partielt processerede transcripter. Dette blev bekræftet ved hjælp af genomisk hybridisering og ved hjælp af sekventering tværs gennem exon-  
10 intron grænserne for hver klon. T11 indeholdt også yderligere sekvens hver ende. T16-4.5 indeholdt en lille insertion nær grænsen mellem exoner 10 og 11, der ikke svarede til intron sekvens. Klonerne CDLS16A, 11a og 13a fra lunge biblioteket indeholdt også fremmede sekvenser af ukendt oprindelse. Klonen C16-1 indeholdt også en kort insertion svarende til en portion af  $\gamma$ -  
15 transposonen af *E. coli*; dette element blev ikke påvist i de øvrige kloner.

5' klonerne PA3-5 frembragt fra pancreas RNA og TB2-7, frembragt fra T84 RNA under anvendelse af den forankrede PCR teknik har identiske sekvenser bortset fra en enkelt nucleotid forskel i længde ved 5' enden som vist i figur 1. 3' klonen, THZ-4 opnået fra T84 RNA, indeholder 3' sekvensen af transcript'en i overensstemmelse med den genomiske sekvens af denne region.

En kombineret sekvens, der repræsenterer den formodede kodende region af CF genet, blev frembragt fra overlappende cDNA kloner. Eftersom de fleste af cDNA klonerne tilsyneladende var afledt fra uprocesserede transcripter, blev yderligere studier udført for at sikre autenticiteten af den kombinerede sekvens. Hver cDNA klon blev første testet for lokalisering til kromosom 7 ved hjælp af hybridiseringsanalyse med en human-hamster somatisk celle hybrid, der indeholdt et enkelt humant kromosom 7 og ved hjælp af pulset felt gel elektroforese. Fin restriktionsenzym-mapping blev også udført for hver klon. Men overlappende regioner var klart identificerbare for de fleste af klonerne, indeholdt mange regioner med unikke restriktionsmønstre.

For yderligere at karakterisere disse cDNA kloner blev de anvendt som prober i gel hybridiseringseksperimenter med EcoRI- eller HindIII-spaltet human genomisk DNA. Som vist i figur 9 kunne fem til seks forskellige restriktionsfragmenter påvises med 10-1 cDNA'en og et tilsvarende antal af

fragmenter med andre cDNA kloner, hvilket tyder på tilstedeværelsen af multiple exoner for det formodede CF gen. Hybridiseringsstudierne identificerede også cDNA klonerne med uproceserede intron sekvenser, eftersom de viste præferentiel hybridisering til en undergruppe af genomiske DNA fragmenter. For de bekræftede cDNA kloner blev deres korresponderende genomiske DNA segmenter isoleret og exonerne og exon/intron grænserne sekventeret. Som vist i figur 7 blev en total på 24 exoner identificeret. Baseret på denne information og resultaterne af fysiske mapping.eksperimenter blev gen locus'et estimeret til at udspænde 250 kb på kromosom 7.

## 2.6 SEKVENSEN

Figur 1 viser nucleotid-sekvensen for den klonede cDNA, der koder CFTR sammen med den udledte aminosyresekvens. Den første base position svarer til det første nucleotid i 5' ekstension klonen PA3-5, som er ét nucleotid længere end TB2-7. Pile viser position af transcription initiation site ved hjælp af primer ekstension analyse. Nucleotid 6129 er efterfulgt af en poly(dA) område. Positioner af exon junctions er markeret ved hjælp af lodrette linjer. Potentielle membran-udspændende segmenter blev sikret under anvendelse af algoritmen fra Eisenberg et al J. Mol. Biol. 179:125 (1984). Potentielle membran-udspændende segmenter som analyseret og vist i figur 11 er omsluttet i boxe i figur 1. I figur 11 er middel hydropati indeks [Kyte and Doolittle, J. Molec. Biol. 157: 105,(1982)] af 9 rest peptider plottet over for aminosyreantallet. De korresponderende positioner af egenskaber af sekundær struktur forudsagt ifølge Garnier et al, [J. Molec. Bio. 157, 165 (1982)] er vist i det nedre panel. Aminosyrer, der omfatter formodede ATP-bindende folder, er understreget i figur 1. Mulige sites for phosphorylering af proteinkinaserne A (PKA) eller C (PKC) er markeret ved hjælp af henholdsvis åbne og lukkede cirkler. Den åbne trekant er over de 3bp (CTT), som er deleteret i CF (se diskussion nedenfor). cDNA klonerne i figur 1 blev sekventeret ved hjælp af dideoxy kæde termination fremgangsmåden, der anvender <sup>35</sup>S mærkede nucleotider ved hjælp af Dupont Genesis 2000™ automatisk DNA sekvenatoren.

Den kombinerede cDNA sekvens udspænder 6129 basepar, bortset fra poly(A) halen ved enden af den 3' utranslaterede region, og den indeholder en

ORF, der er i stand til at kode et polypeptid på 1480 aminosyrer (figur 1). En ATG (AUG) triplet er til stede ved begyndelsen af denne ORF (baseposition 133 - 135). Eftersom den nucleotid-sekvens, der omgiver denne codon (5'-AGACCAUGCA-3'), har de foreslåede egenskaber for konsensus sekvensen (CC) AGCCAUGG(G) af et eukaryot translation initiation site med en særdeles konserveret A ved -3 positionen, er det særdeles sandsynligt, at denne AUG svarer til den første methionin codon for det formodede polypeptid.

For at opnå den sekvens, der svarer til 5' enden af transcript'en blev et primer-ekstension eksperiment udført som beskrevet tidligere. Som vist i figur 10A kunne et primer ekstension produkt på ca. 216 nucleotider observeres, hvilket tyder på, at 5' enden af transcript'en initierede ca. 60 nucleotider opstrøms for enden af cDNA klon 10-1. En modificeret polymerase kæde reaktion (forankret PCR) blev dernæst anvendt til at fremme kloning af 5'ende sekvenserne (figur 10b). To uafhængige 5' ekstension kloner, én fra pancreas og den anden fra T84 RNA, blev karakteriseret ved hjælp af DNA sekventering og viste sig at afvige med kun 1 base i længde, hvilket viser det mest sandsynlige initiation site for transcript'en som vist i figur 1.

Eftersom de fleste af de initiale cDNA kloner ikke indeholdt en poly A hale indikativ for enden af en mRNA, blev forankret PCR også anvendt til 3' enden af transcript'en (Frohman et al, 1988, supra). Tre 3'ekstension oligonucleotider blev frembragt til den terminale portion af cDNA klonen T16-4.5. Som vist i figur 10c blev der opnået 3 PCR produkter af forskellige størrelser. Alle var konsistente med den fortolkning, at enden af transcript'en var ca. 1,2 kb nedstrøms for HindIII site't ved nucleotid position 5027 (se figur 1). DNA sekvensen afledt fra repræsentative kloner var i overensstemmelse med den for T84 cDNA klonen T12a (se figur 1 og 7) og sekvensen for det korresponderende 2,3 kb EcoRI genomiske fragment.

### 3. MOLEKYLÆRGENETIK FOR CF

#### 3.1 SITES FOR EKSPRESSION

For at visualisere transcript'en for det formodede CF gen blev RNA gel blot hybridisering eksperimenter udført med 10-1 cDNA'en som probe. RNA hybridisering resultaterne er vist i figur 8.

RNA prøver blev fremstillet fra vævsprøver opnået fra kirurgisk patologi eller ved autopsi ifølge fremgangsmåder tidligere beskrevet (A.M. Kimmel, S.L. Berger, eds. Meth. Enzymol. 152, 1987). Formaldehyd-geler blev overført til nylon-membraner (Zetaprobe™ BioRad Lab). Membranerne blev dernæst hybridiseret med DNA prober mærket til høj specifik aktivitet ved random priming fremgangsmåden (A. P. Feinberg og B. Vogelstein, Anal. Biochem. 132, 6, 1983) ifølge tidligere publicerede procedurer (J. Rommens et al, Am. J. Hum. Genet. 43, 645 - 663, 1988). Figur 8 viser hybridisering af cDNA klonen 10-1 til en 6,5 kb transcript i de viste væv. Total RNA (10 µg) af hvert væv og Poly A+ RNA (1 µg) af T84 coloncarcinom cellelinje blev adskilt på en 1 % formaldehyd gel. Positionerne for 28S og 18S rRNA båndene er vist. Pile viser positionen af transcripter. Størrelsesbestemmelse blev etableret ved sammenligning med standard RNA markører (BRL Labs). HL60 er en human promycocyt leukæmi cellelinje, og T84 er en human coloncancer cellelinje.

Analyse afslører et fremtrædende bånd på ca. 6,5 kb i størrelse i T84 celler. Tilsvarende blev stærke hybridiseringssignaler også påvist i pancreas og primære kulturer af celler fra nasale polypper, hvilket tyder på, at den modne mRNA af det formodede CF gen er ca. 6,5 kb. Mindre hybridiseringssignaler, der sandsynligvis repræsenterer nedbrydningsprodukter, blev påvist ved de lavere størrelsesområder, men de varierede mellem forskellige eksperimenter. Identiske resultater blev opnået med andre cDNA kloner som prober. Baseret på hybridiseringsbånd-intensiteten og sammenligning med de, der er påvist for andre transcripter under identiske eksperimentelle betingelser, blev det estimeret, at de formodede CF transcripter udgjorde ca. 0,01 % af total mRNA i T84 celler.

En række af andre væv blev også overvåget ved hjælp af RNA gel blot hybridisering analyse i et forsøg på at korrelere ekspressionsmønsteret for 10-1 genet og patologien for CF. Som vist i figur 8 blev transcripter, alle med identisk størrelse, fundet i lunge, colon, svedkirtler (dyrkede epiteliale celler), placenta, lever og parotiskirtel, men signalintensiteterne i disse væv varierede blandt forskellige præparationer og var generelt svagere, end hvad der var påvist i pancreas'en og de nasale polypper. Intensitet varierede blandt forskellige præparationer, for eksempel blev hybridisering i nyre ikke påvist i præparationen vist i figur 8, men kan skelnes i efterfølgende gentagne assays.

Ingen hybridiseringssignaler kunne skelnes i hjernen eller binyrekirtlen (figur 8), ej heller i hudfibroblast og lymfoblast cellelinjer.

I sammendrag så ekspresion af CF genet ud til at forekomme i mange af de undersøgte væv, med højere niveauer i de væv, der var alvorligt berørt i CF. Mens dette epitelialvæv-specifikke ekspresionsmønster er i god overensstemmelse med sygdomspatologien, er ingen signifikant forskel blevet påvist i mængden eller størrelse af transcripter fra CF og kontrolvæv, konsistent med den antagelse, at CF mutationer er raffinerede ændringer på nucleotidniveauet.

### 3.2 DEN VÆSENTLIGE CF MUTATION

Figur 17 viser DNA sekvensen ved F508 deletionen. Til venstre er den omvendte komplement af sekvensen fra baseposition 1649 - 1664 af den normal sekvens (som afledt fra cDNA klonen T16). Nucleotid-sekvensen er fremvist som output (i arbitrære fluorescensintensitet units, y-akse) plottet over for tid (x-akse) for hver af de 2 fotomultiplier rør (PMT#1 og #2) af et Dupont Genesis 2000™ DNA analysesystem. Den korresponderende nucleotid-sekvens er vist underneden. Til højre er den samme region fra en mutant sekvens (som afledt fra cDNA klonen C16). Dobbeltstrengt plasmid DNA templates blev fremstillet ved hjælp af alkalisk lysering procedure. Fem µg af plasmid DNA og 75 ng af oligonucleotid primer blev anvendt i hver sekventeringsreaktion ifølge den protokol, der er anbefalet af Dupont, bortset fra at annealingen blev udført ved 45 °C i 30 min, og at forlængelse/termination trinnet var i 10 min ved 42 °C. De uinkorporerede fluorescens nucleotider blev fjernet ved præcipitering af DNA sekventeringsreaktionsproduktet med ethanol i tilstedeværelsen af 2,5 M ammoniumacetat ved pH 7,0 og skyllet én gang time med 70 % ethanol. Primeren anvendt til T16-1 sekventeringen blev et specifikt oligonucleotid 5'GTTGGCATGCTTTGATGACGCTTC3', der udspænder baseposition 1708 - 1731, og den for C16-1 var den universale primer SK for Bluescript vektoren (Stratagene). Figur 18 viser også DNA sekvensen omkring F508 deletionen, som bestemt ved hjælp af manuel sekventering. Den normale sekvens fra baseposition 1726 - 1651 (fra cDNA T16-1) er vist ved siden af CF sekvensen (fra cDNA C16-1). Det venstre panel viser sekvenserne fra de kodende strenge opnået med B primeren (5'GTTTTCTGGAT-TATGCCTGGGCAC3')

og det højre panel dem fra den modsatte streng med D primeren (5'GTTGGCATGCTTTGATGACGCTTC3'). Parenteserne viser de tre nucleotider hos den normale, der er fraværende i CF (pilehoveder). Sekventering blev udført som beskrevet i F. Sanger, S. Nicklen, A. R. Coulsen, Proc. Nat. Acad. Sci. U. S. A., 74: 5463 (1977).

For at undersøge den proportion af CF patienter, der bærer denne deletion (F508), blev genomiske DNA prøver fra patienter og deres forældre hver amplificeret med oligonucleotid primere, der flankerer mutationen i en polymerase kæde reaktion og hybridiserede til <sup>32</sup>P-mærkede oligonucleotider specifikke for det normale og de formodede mutante sekvenser (se figur 2). Resultaterne af denne analyse er vist i tabel 2.

TABEL 2  
FORDELING AF CF OG NON-CF(N) KROMOSOMER MED OG UDEN 3 bp DELETIONEN

a.	CF kromosomer	N kromosomer
uden deletionen	69	198
med deletionen	145	0
Total	214	198
b.	CF kromosomer	
	med 3 bp deletionen	uden deletionen
CF-PI	62	24
CF-PS	5	9
Uklassificeret	78	36
Total	145 (68 %)	69 (32 %)

Dataene for CF-PI (pancreatisk insuffisient) og CF-PS (pancreatisk tilstrækkelig) kromosomerne blev afledt fra CF familierne anvendt i vores bindingsanalyse. Disse familier blev oprindeligt selekteret uden viden om PI eller PS; de 15 CF-PS familier, der efterfølgende blev identificeret, var ikke indbefattet som en del af denne beregning. De uklassificerede CF kromosomer blev opnået fra "The DNA Diagnosis Laboratory" ved "The Hospital for Sick Children" i Toronto og for hvilke pancreatiske funktionsdata ikke var tilgængelige.



Det kan ses, at 68 % (145/214) af CF kromosomer i den almindelige patientpopulation havde F508 deletionen (tabel 2). I modsætning hertil havde ingen (0/198) af N kromosomer deletionen (tabel 2;  $\chi^2 = 207$ ,  $p < 10^{-57.5}$ ), hvilket tyder på, at denne sekvensændring er specifik for CF, og at den er den væsentlige mutation, der forårsager sygdommen. Ingen rekombination er blevet påvist mellem F508 deletionen og CF.

Andre sekvensforskelle blev bemærket mellem de normale (T16-4.5) og CF (C1-1/5) cDNA klonerne. Ved baseposition 2629 viste T16-4.5 en C, og C1-1/5 havde en T, hvilket resulterer i en Leu til Phe ændring på aminosyreniveauet. Ved position 4555 var basen G i T16-4.5, men A i C1-1/5 (Val til Met). Disse fund formodes at repræsentere sekvenspolymorfisme. Specifik oligonucleotid hybridisering analyse af patient/familie DNA vil identificere disse som andre mulige mutationer. Yderligere nucleotidforskelle blev observeret i de 3' utranslaterede regioner mellem forskellige cDNA kloner og den genomiske DNA sekvens. Sådanne forskelle i sekvenserne og som det er anerkendt, er andre sekvensmodifikationer mulige; for eksempel, hvilke forskelle skyldes normale sekvenspolymorfismer og kloningsartefacter, idet alle sådanne forskelle er i det væsentlige ækvivalente med den sekvens, som er beskrevet i figur 1 i for af dens funktion og dens kommercielle anvendelser.

De omfattende genetiske og fysiske mapping data har rettet molekylære kloningsstudier mod fokus på et lille segment af DNA på kromosom 7. På grund af manglen på kromosomdeletioner og omflytninger i CF og manglen på et veludviklet funktionelt assay for CF gen produktet, krævede identifikationen af CF genet en detaljeret karakterisering af selve locus'et og sammenligning mellem CF allelerne og de normale (N) alleler. Tilfældige, fænotypisk normale, individer kunne ikke indbefattes som kontroller i sammenligningen på grund af den høje frekvens af symptomløse bærere i populationen. Som en følge deraf var kun forældre til CF patienter, der hver per definition bærer et N og et CF kromosom, egnede til analysen. Endvidere var det, på grund af den stærke allel association observeret mellem CF og nogle af de tæt forbundne DNA markører, nødvendigt at udelukke den mulighed, at sekvensforskelle påvist mellem N og CF var polymorfismer associeret med sygdoms-locus'et.

35

### 3.3 IDENTIFIKATION AF RFLP'er OG FAMILIESTUDIER

5 For at bestemme forholdet mellem hver af DNA segmenterne isoleret fra kromosom walking og jumping eksperimenterne og CF blev restriktionsfragment længde polymorfismer (RFLP'er) identificeret og anvendt til at studere familier, hvor crossover-begivenheder tidligere var blevet påvist mellem CF og andre flankerende DNA markører. Som vist i figur 14 blev en total på 18 RFLP'er påvist i 500 kb regionen; 17 af dem (fra E6 til CE1.0) er listet i tabel 3; nogle af dem svarede til tidligere rapporterede markører.

10

Fem af RFLP'erne, nemlig 10-1X.6, T6/20, H1.3 og CE1.0, blev identificeret med cDNA og genomiske DNA prober afledt fra det formodede CF gen. RFLP dataene er præsenteret i tabel 3, med markører i MET og D7S8 regionerne indbefattet til sammenligning. De fysiske afstande mellem disse markører 15 såvel som deres forhold til MET og D7S8 regionerne er vist i figur 14.

**TABEL 3. RFLP'er ASSOCIERET MED CF GENET**

Probe navn	Enzym	Frag-længde	N(a)	CF-PI(a)	A(b)	*(c)	Reference
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			
metH	TaqI	7,5	45	49	0,35	0,05	White et al, supra
		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				

TABEL 3 (fortsat)

pH131	Hinfl	0,4	81	33	0,73	0,15	J.M. Rommens et al, <u>Am. J. Hum. Genet.</u> 43: 645 (1988)
		0,3	18	47			
W3D1.4	HindIII	20	82	33	0,68	0,13	B. Kerem et al, <u>supra</u>
	10	22	47				
H2.3A	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			
EG1.4	HincII	3,8	31	69	0,89	0,17	
	2,8	56	7				
EG1.4	BglII	20	27	69	0,89	0,18	
	15	62	9				
JG2E1	PstI	7,8	69	10	0,88	0,18	X. Estivill et al <u>supra</u> og B. Kerem et al <u>supra</u>
(KM19)		6,6	30	70			
E2.6/E.9	MspI	11	34	6	0,85	0,14	
		8,5	26	55			
H2.8A	NcoI	25	22	55	0,87	0,18	
		8	52	9			
E4.1	MspI	12	37	8	0,77	0,11	G. Romeo, <u>personlig kommunikation</u>
(Mp6d9)		8,5 + 3,5	38	64			
J44	XbaI	15,3	40	70	0,86	0,13	
		15 + ,3	44	6			
10-1X.6	AccI	6,5	67	15	0,90	0,24	
		3,5 + 3	14	60			
10-1X.6	HaeIII	1,2	14	61	0,91	0,25	
		.6	72	15			
T6/20	MspI	8	56	66	0,51	0,54	
		4,3	21	8			
H1.3	NcoI	2,4	53	8	0,87	0,15	
		1 + 1,4	35	69			

TABEL 3 (fortsat)

CE1.0	Ndel	5,5	81	73	0,41	0,03	
		4,7 + 0,8	8	3			
J32	Sacl	15	21	24	0,17	0,02	M.C. Iannuzzi et al Am. J. Genet. 44:695 (1989)
		6	47	38			
J3.11	Mspl	4,2	36	38	0,29	0,04	B.J. Wainright et al, Nature 318: 384 (1985)
		1,8	62	36			
J29	Pvull	9	26	36	0,36	0,06	M..C. Iannuzzi et al, <i>supra</i>
		6	55	36			

NOTER FOR TABEL 3

5

(a) Antallet af N og CF-PI (CF med pancreatisk insufficiens) kromosomer blev afledt fra forældrene i de familier, der blev anvendt i bindingsanalyse [Tsui et al, Cold Spring Harbor Symp. Quant. Biol. 51: 325 (1986)].

10

(b) Standardiseret association (A), der er mindre under indflydelse af fluktuationen af DNA markør allel fordeling blandt N kromosomerne, anvendes her til sammenligningen af Yule's associationskoefficient

$$A = (ad-bc)/(ad+bc),$$

15

hvor a, b, c og d er henholdsvis antallet af N kromosomer med DNA markør allel 1, CF med 1, N med 2 og CF med 2. Relativ risiko kan beregnes under anvendelse af forholdet

20

$$RR = (1+A)/(1-A) \text{ eller dets omvendte.}$$

(c) Allel association (\*), beregnet ifølge A. Chakravarti et al, Am. J. Hum. Genet. 36: 1239, (1984), under antagelse af at frekvensen på 0,02 for CF kromosomer i populationen er indbefattet til sammenligning.

25

På grund af det lille antal af rekombinante familier tilgængelig for analysen, som var forventet fra den tætte afstand mellem de undersøgte markører og

CF, og muligheden for fejl diagnose, var alternative approaches nødvendige til yderligere fin mapping af CF genet.

### 3.4 ALLEL ASSOCIATION

5

Allel association (bindings-ulige vægt) er blevet påvist for mange tæt forbundne DNA markører. Mens nytten ved at anvende allel association til at måle genetisk afstand er uvis, er en samlet korrelation blevet observeret mellem CF og de flankerende DNA markører. En stærk association med CF blev noteret for de tættere DNA markører, D7S23 og D7S122, mens ringe eller ingen association blev påvist for de mere fjerne markører MET, D7S8 eller D7S424 (se figur 1).

10

Som vist i tabel 3 steg graden af association mellem DNA markører og CF (som målt ved hjælp af Yule's associationskoefficient) fra 0,35 for meth og 0,17 for J32 til 0,91 for 10-1X.6 (kun CF-PI patient familier blev anvendt i analysen, eftersom de så ud til at være genetisk mere homogene end CF-PS). Associationskoefficienterne så ud til at være ret konstant over de 300 kb fra EG1.4 til H1.3; fluktuationen påviste, at adskillige lokaliseringer, mest bemærkelsesværdigt ved H2.3A, E4.1 og T6/20, sandsynligvis var på grund af variationen i den allele fordeling blandt N kromosomerne (se tabel 2). Disse data er derfor konsistente med resultatet fra studiet af rekombinante familier (se figur 14). En lignende konklusion kunne også være gøres ved inspektion af de eksterede DNA markør haplotyper associeret med CF kromosomerne (se nedenfor). Imidlertid tillod den stærke allele association påvist over den store fysiske afstand mellem EG1.4 og H1.3 ikke yderligere raffineret mapping af CF genet. Eftersom J44 var den sidste genomiske DNA klon isoleret ved hjælp af kromosom walking og jumping, før en cDNA klon blev identificeret, tilskyndede den stærke allele association påvist for JG2E1-J44 intervallet os til at søge for kandidat gen sekvenser over hele dette interval. Det er af interesse at bemærke, at den højeste grad af allel association i realiteten var påvist mellem CF og 2 RFLP'erne påvist ved hjælp af 10-1X.6, en region nær den større CF mutation.

15

20

25

30

35

Tabel 4 viser parvis allel association mellem DNA markører tæt forbundet med CF. Det gennemsnitlige antal af kromosomer anvendt i disse beregninger var 75 - 80 og kun kromosomer fra CF-PT familier blev anvendt til at score CF

kromosomer. Lignende resultater blev opnået, når Yule's standardiserede association (A) blev anvendt).

5  
10

TABEL 3

N kromosomer

TABLE 3

	ma1D	ma1H	E6	E7	ptH131	D1.4	H2.3A	EG1.4	JG2E1	E2.6	H2.8	E4.1	J44	10-1X.6	T6/20	H1.3	CE1.0	J32	J3.11	J29		
ma1D Banl	0.35	0.49	0.04	0.04	0.05	0.07	0.27	0.06	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
ma1D 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	0.06	0.13	0.15	0.09	0.09	0.05
ma1H TaqI	0.81	0.14	0.01	0.05	0.06	0.06	0.24	0.05	0.08	0.07	0.13	0.15	0.07	0.04	0.02	0.02	0.07	0.02	0.03	0.21	0.04	0.18
E6 TaqI	0.11	0.30	0.00	0.00	0.07	0.06	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 HindI	0.45	0.28	0.23	0.38	0.40	0.91	0.12	0.04	0.08	0.05	0.06	0.03	0.03	0.08	0.16	0.15	0.20	0.04	0.03	0.06	0.08	0.06
W3D1.4 HindIII	0.45	0.28	0.23	0.45	0.47	0.95	0.21	0.02	0.03	0.01	0.06	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 HindII	0.11	0.08	0.07	0.08	0.07	0.20	0.24	0.95	0.87	0.76	0.86	0.81	0.60	0.07	0.13	0.61	0.56	0.04	0.24	0.14	0.15	0.15
EG1.4 BglI	0.03	0.08	0.07	0.08	0.07	0.27	0.40	1.00	0.92	0.77	0.93	0.71	0.55	0.08	0.07	0.56	0.55	0.12	0.20	0.24	0.20	0.20
JG2E1 PstI	0.07	0.08	0.03	0.09	0.06	0.30	0.45	0.93	0.94	0.84	1.00	0.76	0.64	0.11	0.11	0.61	0.57	0.13	0.31	0.26	0.22	0.22
E2.6/E.9 MspI	0.22	0.06	0.07	0.02	0.03	0.20	0.34	0.81	0.82	0.92	0.63	0.97	0.78	0.56	0.52	0.47	0.70	0.32	0.31	0.25	0.22	0.22
H2.8 NcoI	0.05	0.07	0.01	0.06	0.06	0.31	0.31	0.45	0.82	0.93	1.00	0.92	0.74	0.65	0.13	0.18	0.60	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	0.82	0.86	0.94	1.00	0.93	0.71	0.49	0.49	0.49	0.68	0.35	0.27	0.25	0.21	0.21
J44 XbaI	0.18	0.05	0.06	0.01	0.01	0.28	0.26	0.45	0.71	0.69	0.90	0.80	0.85	0.33	0.40	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	0.56	0.64	0.70	0.69	0.59	0.91	0.19	0.36	0.56	0.00	0.02	0.03	0.03
10-1X.6 HaeIII	0.16	0.10	0.25	0.08	0.11	0.41	0.41	0.65	0.64	0.58	0.64	0.70	0.69	0.69	1.00	0.18	0.43	0.62	0.02	0.02	0.08	0.08
T6/20 MspI	0.27	0.07	0.36	0.13	0.13	0.23	0.23	0.29	0.05	0.00	0.01	0.07	0.02	0.01	0.11	0.69	0.69	0.56	0.03	0.21	0.18	0.25
H1.3 NcoI	0.08	0.06	0.06	0.03	0.01	0.30	0.30	0.55	0.71	0.78	0.87	0.90	0.87	0.93	0.92	0.64	0.64	0.12	0.40	0.19	0.13	0.20
CE1.0 NdeI	0.00	0.04	0.02	0.11	0.11	0.25	0.25	0.08	0.69	0.59	0.55	0.43	0.55	0.37	0.44	0.24	0.24	0.07	0.40	0.19	0.20	0.14
J32 SacI	0.03	0.13	0.07	0.17	0.13	0.17	0.24	0.07	0.21	0.21	0.24	0.22	0.21	0.21	0.27	0.28	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	0.12	0.11	0.10	0.13	0.18	0.19	0.15	0.20	0.28	0.24	0.14	0.07	0.81	0.81	0.71	0.71
J29 PvuII	0.11	0.12	0.09	0.10	0.10	0.00	0.00	0.09	0.10	0.10	0.14	0.17	0.20	0.16	0.16	0.29	0.29	0.16	0.06	0.85	0.97	0.97

CF kromosomer

Stærk allel-association blev også påvist blandt subgrupper af RFLP'er på både CF og N kromosomerne. Som vist i tabel 4, så de DNA markører, der er fysisk tæt på hinanden almindeligvis ud til at have stærk association med hinanden. For eksempel blev stærk (i nogle tilfælde næsten komplet) allel association påvist mellem tilgrænsende markører E6 og E7, mellem pH131 og W3D1.4 mellem de Accl og HaeIII polymorfe sites påvist af 10-1X.6 og blandt EG1.4, JG2E1, E2.6(E.9), E2.8 og E4.1. De to grupper af distale markører i MET og D7S8 regionen viste også en vis grad af bindings-uligevægt bland dem selv, men de viste ringe association med markører fra E6 til CE1.0, konsistent med de fjerne lokaliseringer for MET og D7S8. På den anden side kan manglen på association mellem DNA markører, der er fysisk tætte, vise tilstedeværelsen af rekombination hot spots. Eksempler på disse potentielle hot spots er regionen mellem E7 og pH131, omkring H2.3A, mellem J44 og regionerne dækket af proberne 10-1X.6 og T6/20 (se figur 14). Disse regioner, der indeholder hyppige rekombinationsbreakpoints, var anvendelige i den efterfølgende analyse af ekstenderede haplotype data for CF regionen.

### 3.5 HAPLOTYPE ANALYSE

Ekstenderede haplotyper baseret på 23 DNA markører blev frembragt for CF og N kromosomerne i samlingen af familier tidligere anvendt til bindingsanalyse. Under antagelse af rekombination mellem kromosomer af forskellige haplotyper, var det muligt at konstruere adskillige linjer af de observerede CF kromosomer, og også at forudsige lokaliseringen af sygdomslocus'et.

For at opnå yderligere information anvendelig for forståelse af beskaffenheden af forskellige CF mutationer, blev F508 deletionsdataene korreleret med de ekstenderede DNA markør haplotyper. Som vist i tabel 5 kunne fem major grupper af N og CF haplotyper defineres ved hjælp af RFLP'erne inden for eller umiddelbart grænsende op til det formodede CF gen (regionerne 6-8).



TABEL 5 DNA MARKØR HAPLOTYPER, DER UDSPÆNDER CF LOCUS'ET

I. (a)	CF(b)			
	PI (F508)	P8 (F508)	PI others	PS others
1	A A A A A A A B B B B B A A A B B E D D B C A D A D B A A A A A			
2	A A A A A A A A A A A A A B D G B C B B B A A D C D D B B B B C C			
3	A A A A A . A . A . A A A A A A A A A A B . . A A A . . A . B A .			
4	A C			
5	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 .			
6	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			
7	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			
8	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			
9	A A A A B B B C A A A . A A A A A A A A A B B B B B B B B B B B			
10	A A			
11	A A			
12	A A			
13	A A			
14	A A			
15	A A			
16	A A			
17	A A			
18	A A			
19	A A			
20	A A			
21	A A			
22	A A			
23	A A			
24	A A			
25	A A			
26	A A			
27	A A			
28	A A			
29	A A			
30	A A			
31	A A			
32	A A			
33	A A			
34	A A			
35	A A			
36	A A			
37	A A			
38	A A			
39	A A			
40	A A			
41	A A			
42	A A			
43	A A			
44	A A			
45	A A			
46	A A			
47	A A			
48	A A			
49	A A			
50	A A			







	A D	H B	B B	A B	B B	C C	B B	C C	B B	C C	B B	0	0	0	0	1	1	
V. (a)	B A B B B D	C C B C C .	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	C C C C C C	A A A A A A	C C C C C C	A A A A A A	A A A A A A	1	1	1	1	1	1
(b)	B B	C C	A A	B B	C C	C C	A A	C C	C C	A A	C C	A A	0	0	0	1	1	

Andre	B B B B B A G	C C C C C C B	B B B A B .	A A K B B A A	A A B B B B A	B D A E F C B/C	B B B .	A A D C C D A/D	B B A .	B B A .	B B A .	0	0	0	0	1	1
Uklassificeret	.	.	.	.	.	.	.	.	.	.	.	4	10	2	18	6	
<b>Total:</b>											<b>62</b>	<b>15</b>	<b>24</b>	<b>27</b>	<b>98</b>		

TABEL 5 (fortsat)

(a) De ekstenderede haplotype data er afledt fra CF familierne anvendt i tidligere bindingsstudier (se fodnote (a) fra tabel 3) med yderligere CF-PS familier opsamlet efterfølgende (Kerem et al, Am. J. Genet. 44: 827 (1989)). Dataene er vist i grupper (regioner) for at reducere plads. Regionerne er tildelt primært ifølge parvise associationsdata vist i tabel 4 med regioner 6-8, der udspænder den formodede CF locus (F508) deletion, er mellem regionerne 6 og 7). En dash (tankestreg) (-) er vist ved den region, hvor haplotypen ikke er blevet bestemt på grund af inkomplette data eller manglende evne til at fastsætte fase.

Alternative haplotype-tildelinger er også givet, hvor data er inkomplette. Uklassificeret indbefatter de kromosomer, der har mere end 3 ukendte tildelinger. Haplotype-definitionerne for hver af de 9 regioner er:

Region 1-	MetD Bnl	metD TaqI	MetH TaqI
A =	1	1	1
B =	2	1	2
C =	1	1	2
D =	2	2	1
E =	1	2	-
F =	2	1	1
G =	2	2	2

Region 2-	E6 TaqI	E7 TaqI	pH131 HinfI	W3D1.4 HindIII
A =	1	2	2	2
B =	2	1	1	1
C =	1	2	1	1
2 =	2	1	2	2
2 =	2	2	2	1
2 =	2	2	1	1
1 =	1	2	1	2
1 =	1	1	2	2

Region 3-	H2.3A TagI
A =	1
B =	2

Region 4-	EG1.4 HincII	EG1.4 BglI	JG2EI PstI
A =	1	1	2
B =	2	2	1
C =	2	2	2
D =	1	1	1
E =	1	2	1

TABEL 5 (fortsat)

Region 5-	E2.6 MspI	E2.8 NcoI	E4.1 MspI
A =	2	1	2
B =	1	2	1
C =	2	2	2

Region 6-	J44 XbaI	10-1X.610 AccI	-1X.6 HaeIII
A =	1	2	1
B =	2	1	2
C =	1	1	2
D =	1	2	2
F =	2	2	1

5

Region 7-	T6/20 MspI
A =	1
B =	2

Region 8-	H1.3 NcoI	CE 1.0 NdeI
A =	2	1
B =	1	2
C =	1	1
D =	2	2

Region 9-	J32 SacI	J3.11 MspI	J29 PvuII
A =	1	1	1
B =	2	2	2
C =	2	1	2
D =	2	2	1
E =	2	1	1

(b)

5

Antal af kromosomer scoret i hver klasse::

CF-PI(F) = CF kromosomer fra CF-PI patienter med F508 deletionen;

10

CF-PS(F) = CF kromosomer fra CF-PS patienter med F508 deletionen;

CF-PI = Andre CF kromosomer fra CF-PI patienter;

15

CF-PS = Andre CF kromosomer fra CF-PS patienter;

N = Normale kromosomer afledt fra bærer-forældre

20

Det var tydeligt, at de fleste rekombinationer mellem haplotyper forekom mellem regionerne 1 og 2 og mellem regioner 8 og 9, igen i god overensstemmelse med den relativt lange fysiske afstand mellem disse regioner. Andre, mindre hyppige, breakpoints blev bemærket mellem kort afstand intervaller og de svarede almindeligvis til de hotte spots identificeret ved parvis allel association studier som vist ovenfor. Det slående resultat var,



at F508 deletionen associerede næsten udelukkende med Gruppe I, den mest hyppige CF haplotype, hvilket understøtter den position, at denne deletion udgør den større mutation i CF. Mere betydningsfuldt, mens F508 deletionen blev påvist i 89 % (62/70) af CF kromosomerne med AA haplotypen (svarende til de to regioner, 6 og 7) flankerede deletionen, blev ingen fundet i de 14 N kromosomer inden for den samme gruppe ( $\chi^2 = 47,3$ ,  $p < 10^{-4}$ ). F508 deletionen var derfor ikke en almindelig sekvens polymorfisme associeret med kernen af Gruppe I haplotypen (se tabel 5).

Ét af CF kromosomerne, påvist ved hjælp af den specifikke oligonucleotideprobe for F508 deletionen, viste sig at tilhøre en anderledes haplotype gruppe (Gruppe III). Ingen af de 9 andre CF kromosomer eller 17 N kromosomer med den samme gruppe hybridiserede til proben. Dette specifikke hybridiseringsresultat tyder på, at den mutation, der er beliggende på dette kromosom er svarende til F508. Selv om rekombination eller gen omdannelse er mulige mekanismer til at forklare tilstedeværelsen af denne deletion på en non-Gruppe I haplotype, er det mere sandsynligt, at disse 2 Gruppe III kromosomer repræsenterer en tilbagevendende mutationsbegivenhed, en situation svarende til BS og BE mutationerne ved  $\beta$  globin locus'et.

Tilsammen understøtter resultaterne af oligonucleotide-hybridiseringsstudiet og haplotypeanalysen det faktum, at gen locus'et beskrevet her er CF genet, og at 3 bp (F508) deletionen er den mest almindelige i CF.

### 3.6 ANDRE CF MUTATIONER

Associationen af F508 deletionen med 1 almindelig og 1 sjælden CF haplotype tilvejebragte yderligere indsigt i antallet af mutationsbegivenheder, der kunne bidrage til den foreliggende patientpopulation. Baseret på de omfattende haplotype-data, bærer de 2 oprindelige kromosomer, hvor F508 deletionen forekom, sandsynligvis haplotypen - AAAAAA- (Gruppe Ia) og - CBAACBA- (Gruppe IIIa), som defineret i tabel 5. De øvrige Gruppe I CF kromosomer, der bærer deletionen, er sandsynligvis rekombinationsprodukter afledt fra det oprindelige kromosom. Hvis CF kromosomerne i hver haplotype gruppe anses for at være afledt fra den samme oprindelse, ville kun 3-4 yderligere mutationsbegivenheder være forudsagt (se tabel 5). Eftersom imidlertid mange

af CF kromosomerne i den samme gruppe er markant forskellige fra hinanden, er yderligere underinddeling inden for hver gruppe mulig. Som én følge deraf kunne et højere antal af uafhængige mutationsbegivenheder betragtes, og dataene tyder på, at mindst 7 yderligere, formodede mutationer også bidrager til CF-PI fænotypen (se tabel 4). De mutationer, der fører til CF-PS subgruppen, er sandsynligvis mere heterogene.

De 7 yderligere CF-PI mutationer er repræsenteret af haplotyperne: -CAAAAAA- (Gruppe Ib), -CABCAAD- (Gruppe Ic), -BBBAC- (Gruppella), -CABBBAB- (Gruppe Va). Selv om den molekylære defekt i hver af disse mutationer endnu ikke er blevet defineret, er det klart, at ingen af disse mutationer alvorligt påvirker regionen svarende til de oligonucleotid-bindende sites, der er anvendt i PCR/hybridisering eksperimentet.

### 15 3.7 PANCREATISK SUFFICIENS

CF-PS er defineret klinisk som tilstrækkelig pancreatisk eksokrine funktion til spaltning af føde; imidlertid varierer niveauet for residual pancreatisk enzymaktivitet i fordøjelsessystemet fra patient til patient. Tidligere haplotype-data tydede på, at CF-PI og CF-PS patienterne skyldes forskellige mutante alleler. Selv om den basale biokemiske defekt i CF endnu ikke er blevet defineret, er det muligt, at den residuale pancreatiske enzym-aktivitet hos CF-PS patienter er en direkte afspejling af aktiviteten af det mutante CF genprodukt. Således ville den residuale eksokrine funktion bibragt af en mild (CF-PS) allel, selv om den er meget lavere end den for det normale genprodukt, udgøre en dominant fænotype i forhold til den for mere alvorlige (CF-PI) mutationer med ringe eller ingen funktion. Det følger deraf, at kun patienter, der bærer 2 kopier af alvorlige alleler, ville være CF-PI, og at patienter, der bærer 1 eller 2 milde alleler, ville være CF-PS.

30 For at teste den ovennævnte hypotese, kunne informationen om proportionen af CF patienter, der bærer F508 deletionen, anvendes. Under antagelse af at en alvorlig mutation er recessiv for en mild mutation og en fordeling af CF alleler blandt patientpopulationen ifølge Hardy-Weinberg loven kunne frekvensen af alvorlige alleler estimeres til at være 0,92 og den for de milde alleler (M), 0,08 (se tabel 6).

TABEL 6  
POPULATIONSANALYSE AF CF-PI OG CF-PS

	Antaget genotype(a)	Forudsagt frekvens (b)	Observeret (c)	Forventet (d)
Pancreatisk insufficient (PI)	FF	0,459	21	21,1
	FS	0,331	14	15,2
	SS	0,060	4	2,7
	Total	0,850	39	-
Pancreatisk sufficient (PS)	FM	0,106	15 (e)	14,8
	SM	0,038	6	6,2
	MM	0,006		
	Total	0,150	21	

5 (a) Allel-betegnelser: F = 3 bp deletionen (deletion af phenylalanin ved aminosyreposition 508); S = ukarakteriseret alvorlige mutante alleler; M = ukarakteriseret milde mutante alleler.

10 (b) Under antagelse af at den CF-PI mutante fænotype er recessiv for den CF-PS mutante fænotype, kunne frekvensen af CF-PI mutante alleler, herunder 3 bp deletionen, estimeres fra den observerede proportion af CF-PI patienterne i CF klinikken [Corey et al *J. Pediatr.* 115:274 (1989)], i.e.  $(0,85) = 0,92$ . Den observerede allel-frekvens for F i den totale CF population er 0,68 (tabel 3); frekvensen for S =  $0,92 - 0,68 = 0,24$ ; Frekvensen for M =  $1 - 0,92 = 0,08$ . Frekvensen for hver genotype blev dernæst beregnet ved at anvende af Hardy-Weinberg loven.

20 (c) Antallet af CF-PI og CF-PS patienter i hver kategori blev opnået ved hjælp af oligonucleotid hybridisering analyse som illustreret i figur 15. Patienterne var fra CF familierne anvendt i vores bindingsanalyse med 14 yderligere CF-PS patienter/familier fra et efterfølgende studium. Eftersom SM og MM ikke kunne skelnes genotypisk eller fænotypisk, blev de kombineret i analysen.

(d) Det forventede antal blev beregnet for CF-PI og CF-PS efter normalisering inden for hver gruppe.  $\chi^2$  for fit er 0,86, d.f. = 3,  $0,74 < p < 0,90$ .

5 (e) Dette antal er højere, end hvad der ville forventes (15 observeret vs. 9,6 forventet), hvis F508 deletionen er i Hardy-Weinberg ligevægt blandt alle CF kromosomer ( $\chi^2 = 6,48$ , d.f. = 1,  $p < 0,011$ ).

10 Eftersom størstedelen af CF-PI patienter viste sig at være homozygote for F508 mutationen (F), var det rimeligt at antage, at denne mutation svarede til én af de alvorlige alleler. Givet den observerede frekvens af F (0,68) i den undersøgte CF population kunne frekvensen af de resterende alvorlige alleler (S) afledes. Proportionen af FF, SS, MM, FS, FM og SM patienter blev dernæst beregnet. Eftersom individer med SM og MM ikke kunne skelnes fænotypisk eller genotypisk, blev de kombineret i analysen. Som vist i tabel 6  
15 var de observerede frekvenser for alle 5 grupper af patienter som forventet fra denne hypotese.

Den ovennævnte analyse tilvejebringer således stærk støtte for vores position, at CF-PI skyldes af tilstedeværelsen af 2 alvorlige alleler, og at en CF-PS patient bærer enten en enkelt alvorlig allel eller 2 milde alleler. Denne model  
20 forklarer også den lavere frekvens af F508 deletionen i CF-PS populationen end i CF-PI populationen og det overskydende antal af CF-PS patienter med én kopi af deletionen (se note i tabel 6).

25 Givet den forudsagte dominante fænotype bibragt af M allelerne var det nødvendige at undersøge CF kromosomerne hos CF-PS patienter individuelt med henblik på at identificere de, der bærer M allelerne. Som vist i tabel 7 bærer fem af de 7 repræsentative CF-PS patienter én kopi af F508 deletionen; mindst 5 forskellige haplotyper kunne tilskrives de øvrige CF kromosomer.  
30

Tabel 7. Haplotyper af CF kromosomer i CF-PS individer og familier med MI

Familie #	1	2	3	4	5	6	7	8	9	CF alleler
-----------	---	---	---	---	---	---	---	---	---	------------

## (a) CF-PS individer

5

3	A	A	A	A	A	A	-	-	A	F (Gruppe Ia
	D	C	B	A	A	C	B	A	A	M (forudsagt, Gruppe IIIb)

14	B	A	A	A	-	A	A	A	A	F (Gruppe Ia
	B	C	B	B	-	B	A	C	C	M (forudsagt, Gruppe IIa)

27	A	B	-	A	A	A	A	A	E	F (Gruppe Ia
	A	C	-	A	A	A	A	A	A	M (forudsagt, Gruppe Ib)

29	A	C	-	C	-	A	A	A	B	F (Gruppe Ia
	B	A	-	B	-	B	A	A/C	B	M (forudsagt, Gruppe IIa)

10

40	D	A	A	A	A	A	A	A	B	F (Gruppe Ia
	F	C	B	A	A	C	B	C	A	M (forudsagt, Gruppe IV)

51	C	C	A	B	B	B/C	A	C	A	M (forudsagt, Gruppe Iia)
	F	D	A	B	B	B/C	A	C	C	

54	B	C	A	B	C	C	A	C	A	M eller S (forudsagt, Gruppe Vb)
	B	B	A	A	A	C	B	A	A	M (forudsagt, Gruppe IIIb)

## (b) Familier med MI

4	B	A	A	A	A	A	A	A	A	F (Gruppe Ia)
	B	A	A	A	A	A	A	A	A	F (Gruppe Ia)

10	D	B	A	A	-	A	-	A	A	F (Gruppe Ia)
	A	D	A	A	-	A	A	A	B	F (Gruppe Ia)

23	A	E	B	A	A	A	A	A	E	F (Gruppe Ia)
	B	C	A	A	A	A	A	A	B	S (forudsagt, Gruppe Ib)

5

28	A	A	-	A	A	A	A	A	C	F (Gruppe Ia)
	A	A	-	A	A	A	A	A	B	F (Gruppe Ia)

33	B	B	-	A	A	A	-	A	B	F (Gruppe Ia)
	B	A	-	A	A	A	A	A	B	F (Gruppe Ia)

49	A	A	A	A	A	A	A	A		F (Gruppe Ia)
										F (Gruppe Ia)

10

(a) Haplotype-definitionerne er de samme som i tabel 5.

(b) Allel-betegnelser er de samme som i tabel 6:

F = F508 deletionen;

S = ukarakteriseret alvorlig mutant allel;

M = ukarakteriseret mild mutant allel.

- 5 Disse sidstnævnte observationer tilvejebringe yderligere støtte for, at størstedelen af CF-PS patienter er sammensatte heterozygote.

#### 4.0 CFTR PROTEIN

- 10 Som drøftet med hensyn til DNA sekvensen i figur 1 forudsagde analyse af sekvensen af de overlappende cDNA kloner et uprocesseret polypeptid på 1480 aminosyrer med en molekylmasse på 168.138 dalton. Som det senere beskrives, kan proteinets molekylvægt på grund af polymorfismer i proteinet variere på grund af mulige substitutioner eller deletion af visse aminosyrer.
- 15 Molekylvægten vil også ændres på grund af additionen af kulhydratenheder til dannelse af et glycoprotein. Det skal også forstås, at det funktionelle protein i cellen vil svare til det uprocesserede polypeptid, men kan være modificeret på grund af cellemetabolisme.

- 20 Den foreliggende opfindelse tilvejebringer således oprenset normal CFTR polypeptid kendetegnet ved en molekylvægt på omtrentligt 170.000 dalton og med epitelial celle transmembran ion konduktans aktivitet. Det normale CFTR polypeptid, der i det væsentlige er fri af andre human proteiner, er kodet af de ovenfor nævnte DNA sekvenser og ifølge én udførelsesform, den i figur 1. Et
- 25 sådant polypeptid udviser den immunologiske eller biologiske aktivitet af normal CFTR polypeptid. Som det senere vil blive drøftet, kan CFTR polypeptidet og fragmenterne deraf fremstilles ved hjælp af kemisk eller enzymatisk peptidsyntese eller udtrykkes i et passende dyrket cellesystem. Den foreliggende opfindelse tilvejebringer også oprenset mutant CFTR
- 30 polypeptid, der er kendetegnet ved cystisk fibrose associeret aktivitet i humane epiteliale celler. Et sådant mutant CFTR polypeptid, som i det væsentlige er fri af andre human proteiner, kan kodes af den mutante DNA sekvens.

#### 4.1 STRUKTUR FOR CFTR

35

Den mest karakteristiske egenskab ved det forudsagte protein er tilstedeværelsen af to gentagne motiver, hvoraf hver består af et sæt af

aminosyrerester, der er i stand til at udspænde membranen adskillige gange efterfulgt af sekvens-lignende konsensus nucleotid (ATP)-bindende folder (NBF'er) (figur 11, 12 og 16). Disse karakteristika ligner bemærkelsesværdigt de for det mammale multilægemedel-resistente P-glycoprotein og en række af andre membran-associerede proteiner, hvilket således indebærer, at det forudsagte CF genprodukt sandsynligvis er involveret i transporten af stoffer (ioner) tværs gennem membranen og sandsynligvis er et medlem af en membranprotein super familie.

Figur 13 er en skematisk model af det forudsagte CFTR protein. I figur 13 viser cylindere membran-udspændende helices, skraverede cirkler viser NBF'er. Den stiplede cirkel er det polære R-domæne. De 6 membran-udspændende helices i hver halvdel af molekylet er vist som cylindere. De indre cytoplasmatiske orienterede NBF'er er vist som skraverede kugler med sprækker til at vise midlerne for indgang af nucleotidet. Det store polære R-domæne, der forbinder de to halvdele, er repræsenteret af en stiplet kugle. Ladede individuelle aminosyrer inden for de transmembrane segmenter og på R-domæne overfladen er vist som små cirkler, der indeholder det ladede tegn. Nettoladninger på de interne og eksterne loops, der forbinder de membrane cylindere og på regioner af NBF'erne er indeholdt i åbne firkanter. Sites for phosphorylering af proteinkinaserne A eller C er vist ved hjælp af henholdsvis lukkede og åbne trekkanter. K, R, H, D og E er standard nomenklatur for aminosyrerne, henholdsvis lysin, arginin, histidin, asparaginsyre og glutaminsyre.

Hver af de forudsagte membran-associerede regioner i CFTR proteinet består af 6 stærkt hydrofobe segmenter, der er i stand til at af udspænde et lipid-dobbeltlag ifølge algoritmerne af Kyte og Doolittle og af Garnier et al (J. Mol. Biol. 120, 97 (1978) (figur 13). De membran-associerede regioner er hver efterfulgt af en stor hydrofil region, der indeholder NBF'erne. Baseret på sekvensopstilling med andre kendte nucleotid-bindende proteiner, omfatter hver af de formodede NBF'er i CFTR mindst 150 rester (figur 13). 3 bp deletionen påvist i størstedelen af CF patienter er lokaliseret mellem de 2 mest udpræget konserverede segmenter af den første NBF i CFTR. Aminosyresekvensidentiteten mellem den region, der omgiver phenylalanin-deletionen og de korresponderende regioner fra en række af andre proteiner tyder på, at denne region er af funktionel betydning (figur 16). En hydrofob



aminosyre, sædvanligvis én med en aromatisk sidekæde, er til stede i de fleste af disse proteiner ved den position, der svarer til F508 af CFTR proteinet. Det forstås, at aminosyrepolymorfismer kan eksistere som en følge af DNA polymorfismer.

5

Figur 16 viser opstilling af de 3 mest konserverede segmenter af de ekstenderede NBF'er af CFTR med sammenlignelige regioner af andre proteiner. Disse 3 segmenter består af rester 433-473, 488-513 og 542-584 af den N-terminale halvdel og 1219-1259, 1277-1302 og 1340-1382 af den C-terminale halvdel af CFTR. De tunge overliggende udpeger regionerne med størst lighed. Yderligere generel homologi kan ses selv uden indføringen af gaps.

10

På trods af den overordnede symmetri i proteinets struktur og sekvenskonserveringen af NBF'erne er sekvenshomologi mellem de to halvdele af det forudsagte CFTR protein beskeden. Dette er påvist i figur 12, hvor aminosyrerne 1 - 1480 er repræsenteret på hver akse. Linjer på hver side af identitet diagonalen viser positionerne for interne ligheder. Mens derfor fire sæt af intern sekvensidentitet kan påvises som vist i figur 12, under anvendelse af Dayhoff scoringsmatrix'en som anvendt af Lawrence et al. [C. B. Lawrence, D. A. Goldman, og R. T. Hood, Bull Math Biol. 48, 569 (1986)], er kun tre af disse synlige ved lav tærskelværdi opsætninger for standard deviation. Den stærkeste identitet er mellem sekvenser ved carboxylenderne af NBF'erne. Af de 66 opstillede rester er 27 % identiske og yderligere 11 % er funktionelt ens. Den samlede svage interne homologi er i modsætning til den meget højere grad ( $> 70\%$ ) i P-glycoprotein, for hvilket en gen duplikation hypotese er blevet foreslået (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). Manglen på konservering i de relative positioner af exon-intron grænserne kan være argument mod en sådan model for CFTR (figur 2).

20

25

30

Eftersom der tilsyneladende ikke er nogen signal-peptid sekvens ved aminoterminalen af CFTR, er det særdeles ladede hydrofile segment, der forudgår den første transmembrane sekvens, sandsynligvis orienteret i cytoplasmaet. Hver af de 2 sæt af hydrofobe helices forventes at danne 3 tværgående loops tværs gennem membranen, og kun lidt sekvens af hele proteinet forventes at være eksponeret for den ydre overflade, bortset fra regionen mellem

35

transmembrant segment 7 og 8. Det er af interesse at bemærke, at den sidstnævnte region indeholder to potentielle sites for N-forbundet glycosylering.

5 Hver af de membran-associerede regioner er efterfulgt af en NBF som vist ovenfor. Derudover kan et særdeles ladet cytoplasmatisk domæne være identificeret i midten af det forudsagte CFTR polypeptid, der forbinder de 2 halvdele af proteinet. Dette domæne, betegnet R-domænet, er operationelt defineret af en enkelt stor exon, hvor 69 af de 241 aminosyrer er polære rester  
10 arrangeret i skiftende klynger af positive og negative ladninger.

Endvidere er 9 af de 10 konsensus sekvenser påkrævet for phosphorylering by protein kinase A (PKA), og 7 af de potentielle substrat-sites for proteinkinase C (PKC) fundet i CFTR lokaliseret i denne  
15 exon.

#### 4.2 FUNKTION AF CFTR

Egenskaber for CFTR kan afledes ud fra sammenligning med andre membran-associerede proteiner (figur 16). Foruden den samlede strukturelle lighed med  
20 det mammale P-glycoprotein viser hver af de to forudsagte domæner i CFTR også bemærkelsesværdig lighed med enkelt-domæne strukturen af hemolysin B fra *E. coli*, og produktet af White genet fra *Drosophila*. Disse sidstnævnte proteiner er involveret i transporten af det lytiske peptid fra henholdsvis hemolysin-systemet og fra øjepigment-molekyler. Vitamin B12 transport-systemet fra *E. coli*, BtuD og MbpX, der er et liverwort chloroplast gen, hvis funktion er ukendt, har også et lignende strukturelt motiv. Derudover deler  
25 CFTR proteinet strukturel lighed med adskillige af de periplasmatiske opløst-stof transportsystemer fra gram-negative bakterier, hvor den transmembrane region og de ATP-bindende folder er indeholdt i separate proteiner, der fungerer i samarbejde med et tredje substrat-bindende polypeptid.  
30

Det samlede strukturelle arrangement af de transmembrane domæner i CFTR ligner adskillige kation-kanal proteiner og visse kation-translocerende ATPaser  
35 såvel som den for nyligt beskrevne adenylat-cyclase fra bovin hjerne. Den funktionelle betydning af denne topologiske klassificering, bestående af 6 transmembrane domæner, forbliver spekulativ.

Korte regioner af sekvensidentitet er også blevet påvist mellem de formodede transmembrane regioner fra CFTR og andre membran-udspændende proteiner.

5 Det er interessant, at der også er sekvenser, 18 aminosyrer i længde, der er beliggende ca. 50 rester fra carboxylterminalen fra CFTR og af serin/threoninkinase protoonkogenet af *Xenopus laevis*, som er identiske ved 12 af disse positioner.

10 Endelig er en aminosyresekvens-identitet (10/13 konserverede rester) blevet bemærket mellem et hydrofilt segment (position 701-713) inden for det stærkt ladede R-domæne fra CFTR og en region, som umiddelbart forudgår det første transmembrane loop af natrium-kanalerne i både rotte hjerne og ål. Det ladede R-domæne fra CFTR deles ikke med det topologisk tæt beslægtede P-glycoprotein; det 241 aminosyre-forbindende peptid er tilsyneladende den største forskel mellem de to proteiner.

15 I sammendrag viser egenskaber for den primære struktur af CFTR proteinet dets besiddelse af egenskaber egnede for deltagelse i reguleringen og kontrollen af ion-transport i de epitheliale celler af væv berørt ved CF. Sikker tilhæftning til membranen i to regioner fungerer til at positionere dets tre større intracellulære domæner (nucleotid-bindende folder 1 og 2 og R-domænet) nær den cytoplasmatiske overflade af cellemembranen, hvor de kan modulere ion-bevægelse via kanaler dannet enten af selve de CFTR transmembrane segmenter eller af andre membranproteiner.

20 I lyset af de genetiske data, vævsspecificiteten og de forudsagte egenskaber for CFTR proteinet er det rimeligt at konkludere, at CFTR er direkte ansvarlig for CF. Det er imidlertid stadig uklart, hvordan CFTR er involveret i reguleringen af ion-konduktans tværs gennem den apicale membran af epitheliale celler.

30 Det er muligt, at CFTR fungerer som en ion-kanal i sig selv. Som vist i figur 13 indeholder 10 af de 12 transmembrane regioner én eller flere aminosyrer med ladede sidekæder, en egenskab, der svarer til hjerne natrium-kanalen og de GABA receptor chlorid-kanal subunits, hvor ladede rester er til stede i 4 af de 6, og 3 af de 4, respektive membran-associerede domæner per subunit eller

35

gentaget unit. Den amfifatiske beskaffenhed af disse transmembrane segmenter formodes at bidrage til den kanal-dannende kapacitet for disse molekyler. Alternativt kan CFTR ikke være en ion-kanal men i stedet fungere til at regulere ion-kanal aktiviteter. Til støtte for den sidstnævnte antagelse ser ingen af de oprensede polypeptider fra trachea og nyre, der er i stand til at rekonstituere chlorid-kanaler i lipidmembraner, [Landry et al, Science 224:1469 (1989)] ud til at være CFTR, hvis der vurderes på basis af molekyllmassen.

I begge tilfælde tyder tilstedeværelsen af ATP-bindende domæner i CFTR på, at ATP hydrolyse er direkte involveret og påkrævet for transportfunktionen. Den høje tæthed af phosphoryleringssites for PKA og PKC og klyngerne af ladede rester i R-domænet kan begge fungere til at regulere denne aktivitet. Deletionen af en phenylalanin rest i NBF'en kan forhindre rigtig binding af ATP eller den konformationelle ændring, som denne normalt fremkalder og som følge deraf resulterer i den observerede insensitivitet til aktivering af PKA- eller PKC-medieret phosphorylering af den CF apicale chlorid-konduktans pathway. Eftersom det forudsagte protein indeholder adskillige domæner og tilhører til en familie af proteiner, der ofte fungerer som dele af multikomponent molekylære systemer, kan CFTR også deltage i epiteliale vævsfunktioner for aktivitet eller regulering ikke relateret til ion-transport.

Idet det isolerede CF gen (cDNA) nu er tilgængeligt, er det muligt at definere den basale biokemiske defekt ved CF og at yderligere belyse kontrollen af ion-transport pathways i epiteliale celler i almindelighed. Det mest vigtige er, at viden således opnået indtil nu ud fra den forudsagte struktur af CFTR sammen med den yderligere information fra studier af selve proteinet tilvejebringer en basis for udviklingen af forbedrede midler til behandling af sygdommen. I sådanne studier er antistoffer blevet rejst mod CFTR proteinet som beskrevet senere.

30

#### 4.3 PROTEINOPRENSNING

CFTR proteinet kan oprenses ved hjælp af fremgangsmåder selekteret på basis af egenskaber som afsløret af dets sekvens. Eftersom det for eksempel besidder karakteristiske egenskaber for et integralt membran protein, en membran fraktion af de epiteliale celler, hvor udtrykkes kraftigt (e.g. den dyrkede coloncarcinom cellelinje, T84) er først isoleret under anvendelse af

35

etablerede fremgangsmåder [J. E. Langridge, et al, Biochim. Biophys. Acts. 751: 318 (1983)]. De perifere proteiner af disse membraner er de, der fjernes ved hjælp af ekstraktion med høje saltkoncentrationer, højt pH eller chaotrope midler, såsom lithiumdiiodsalicylat. Alle de resterende integrale proteiner, herunder CFTR proteinet, solubiliseres dernæst under anvendelse af et 5 detergens, såsom octylglucosid (Landry, et al, supra), CHAPS [D. J. Beres et al, J. Biol. Chem. 262: 10613(1987)], eller andre forbindelser af tilsvarende virkning. Ved at gøre brug af de nucleotid-bindende domæner i CFTR anvendes cibacron-blå [S. T. Thompson et al. Proc. Nat. Acad. Sci. U. S. A. 10 72: 669 (1975)] affinitetskromatografi dernæst til at binde CFTR proteinet og fjerne det fra andre integrale proteiner fra den detergent-stabiliserede blanding. Eftersom CFTR er et glycoprotein, kan differential lectin-kromatografi frembringe yderligere oprensning [Riordan et al. J. Biol. Chem. 254: 1270 (1979)]. Endelig oprensning til homogenitet opnås dernæst under anvendelse 15 af andre standard protein oprensningsprocedurer; i.e. ionbytningskromatografi, gel-permeationskromatografi, adsorptionskromatografi eller isoelektrisk fokusering, alt efter hvad der er nødvendigt. Alternativt gøres der brug af enkelt trin oprensningsprocedurer, såsom immuno-affinitetskromatografi under anvendelse af immobiliserede antistoffer mod CFTR proteinet (eller fragmenter 20 deraf) eller præparativ polyacrylamidgel elektroforese under anvendelse af avanceret instrumentering, såsom The Applied Biosystems "230A HPEC System". Baseret på erfaring i oprensningen af P-glycoprotein (Riordan et al, supra, et yderligere medlem af det generelle kategori af nucleotid-bindende transport-associerede membranproteiner, fremme oprensningen af CFTR 25 proteinet.

Foruden oprensning fra væv og celler, hvor CFTR proteinet udtrykkes kraftigt, anvendes lignende procedurer til at oprense CFTR fra celler transficeret med vektorer, der indeholder CF genet (cDNA) som beskrevet ovenfor. 30 Proteinprodukter, der er resultat af ekspression af modificeret version af cDNA sekvensen, oprenses på en lignende måde. Kriterier for homogeniteten af således tilvejebragt protein indbefatter de, der er standard for området af proteinkemi, herunder én- og to-dimensional gelelektroforese og N-terminal aminosyrebestemmelse. Det oprensede protein anvendes i yderligere fysisk 35 biokemisk analyse med henblik på at bestemme egenskaber af dets sekundære og tertiære struktur, for at fremme designet af lægemidler med henblik på at fremme den rigtige funktion af de mutante CF former. I

forberedelse til anvendelse i proteinterapi betragtes fraværet af potentielt toksiske forurenende stoffer. Det er anerkendt, at den hydrofobe beskaffenhed af proteinet nødvendiggør indbefatningen af amfifile forbindelser, såsom detergenser og andre [J. V. Ambud Kar og P. C. Baloney J. Biol. Chem. 261: 10079 (1986)] på alle stadier af dets handling.

## 5.0 CF SCREENING

### 5.1 DNA BASERET DIAGNOSE

10 Givet den viden om den større mutation, som er beskrevet heri, kan bærer-screening og prænatal diagnose udføres som følger.

15 Høj-risiko populationen for cystisk fibrose er kaukasiere. For eksempel ville hver kaukasiske kvinde og/eller mand i den fødedygtige alder blive screenet med henblik på at bestemme, om hun eller han var en bærer (ca. en 5 % sandsynlighed for hvert individ). Hvis begge er bærere, er de et par med høj risiko for et barn med cystisk fibrose. Hvert barn af de par, der er i risiko, har en 25 % risiko for at blive ramt af cystisk fibrose. Proceduren for at bestemme bærer-status under anvendelse af prøberne beskrevet heri er som følger.

20 Én væsentlig anvendelse af DNA sekvens informationen om de normale og mutante CF gener er i området for genetisk testning, bærer-påvisning og prænatal diagnose. Individuer, der bærer mutationer i CF genet (sygdomsbærer eller patienter) kan påvises på DNA niveauet med anvendelsen af en række teknikker. Den genomiske DNA anvendt til diagnosen kan opnås fra kropsceller, såsom de, der er til stede present i perifert blod, urin, spyt, vævsbiopsi, kirurgiske prøver og autopsimateriale. DNA'en kan anvendes direkte til påvisning af specifik sekvens eller kan amplificeres enzymatisk in vitro ved at anvende PCR [Saiki et al. Science 230: 1350-1353, (1985), Saiki et al. Nature 324: 163-166 (1986)] forud for analyse. RNA eller dens cDNA form kan også anvendes til det samme formål. Nylige oversigtsartikler af dette emne er blevet præsenteret af Caskey, (Science 236: 1223-8 (1989) og af Landegren et al (Science 242: 229-237 (1989)).

35 Påvisningen af specifik DNA sekvens kan opnås ved hjælp af fremgangsmåder, såsom hybridisering under anvendelse af specifikke oligonucleotider (Wallace et al. Cold Spring Harbour Symp. Quant. Biol. 51:

257 - 261 (1986)], direkte DNA sekventering [Church og Gilbert, Proc. Nat. Acad. Sci. U. S. A. 81: 1991-1995 (1988)], anvendelsen af restriktionsenzymmer (Flavell et al. Cell 15: 25 (1978), Geever et al, Proc. Nat. Acad. Sci. U. S. A. 78: 5081 (1981)], skelnen på basis af elektroforetisk mobilitet i geler med denaturerende reagens (Myers og Maniatis, Cold Spring Harbour Sym. Quaint. Biol. 51: 275 - 284 (1986)], RNase beskyttelse [Myers, R. M., Larin, J., og T. Maniatis Science 230: 1242 (1985)], kemisk spaltning (Cotton et al Proc. Nat. Acad. Sci. U. S. A. 85: 4397 - 4401, (1985)) og den ligase-medierede påvisning procedure [Landegren et al Science 241:1077 (1988)].

10

Oligonucleotider specifikke for normale eller mutante sekvenser syntetiseres kemisk under anvendelse af kommercielt tilgængelige maskiner, mærket radioaktivt med isotoper (såsom <sup>32</sup>P) eller non-radioaktivt (med tags, såsom biotin (Ward og Langer et al. Proc. Nat. Acad. Sci. U. S. A. 78: 6633 - 6657 (1981)), og hybridiseret til individuelle DNA prøver immobiliseret på membraner eller andre faste understøtninger ved hjælp af dot-blot eller overførsel fra geler efter elektroforese. Tilstedeværelsen eller fraværet af disse specifikke sekvenser visualiseres ved hjælp af fremgangsmåder, såsom autoradiografi eller fluorometriske (Landegren et al, 1989, supra) eller kolorimetriske reaktioner (Gebeyehu et a. Nucleic Acids Research 15: 4513 - 4534 (1987)). En udførelsesform af denne oligonucleotid screeningsfremgangsmåde er blevet anvendt til påvisningen af F508 deletionen som beskrevet heri.

15

20

25

30

35

Sekvensforskelle mellem normale og mutanter kan afsløres ved hjælp af den direkte DNA sekventering fremgangsmåde af Church og Gilbert (*supra*). Klonede DNA segmenter kan anvendes som prober til at påvise specifikke DNA segmenter. Sensitiviteten for denne fremgangsmåde er stærkt forøget, når der kombineres med PCR [Wrichnik et al, Nucleic Acids Res. 15: 529-542 (1987); Wong et al, Nature 330: 384 - 386 (1987); Stoflet et al, Science 239: 491 -494 (1988)]. I den sidstnævnte procedure anvendes en sekventering primer, der ligger inden for den amplificerede sekvens, med dobbeltstrenget PCR produkt eller enkeltstrenget template frembragt ved hjælp af en modificeret PCR. Sekvensbestemmelsen udføres ved hjælp af traditionelle procedurer med radioaktivt mærkede nucleotider eller ved hjælp af automatiske sekventeringsprocedurer med fluorescens-tags.

5 Sekvensændringer kan lejlighedsvis frembringe tilfældige restriktionsenzym-  
genkendelses-sites, der afsløres ved at anvende passende enzymspaltning  
efterfulgt af traditionel gel-blot hybridisering (Southern, J. Mol. Biol 98: 503  
(1975)]. DNA fragmenter, der bærer site't (enten normale eller mutante) er  
10 påvist ved deres reduktion i størrelse eller stigninger af korresponderende  
restriktionsfragment-antal. Genomiske DNA prøver kan også amplificeres ved  
hjælp af PCR forud for behandling med det passende restriktionsenzym;  
fragmenter med forskellige størrelser visualiseres dernæst under UV lys i  
tilstedeværelsen af ethidiumbromid efter gel elektroforese.

10 Genetisk testning baseret på DNA sekvens forskelle kan opnås ved påvisning  
af ændring i elektroforetisk mobilitet af DNA fragmenter i geler med eller uden  
denaturerende reagens. Små sekvens-deletioner og -insertioner kan  
visualiseres ved hjælp af høj resolution gel elektroforese. For eksempel er  
15 PCR produktet med 3 bp deletionen klart til at skelne fra den normale sekvens  
på en 8 % non-denaturerende polyacrylamidgel. DNA fragmenter med  
forskellige sekvenssammensætninger kan skelnes på denaturerende  
formamid gradient gel, hvor mobiliteterne af forskellige DNA fragmenter er  
forsinket i gelen ved forskellige positioner ifølge deres specifikke "delvist-  
20 smeltende" temperaturer (Myers, supra). Derudover kan sekvensændringer,  
især små deletioner, påvises som ændringer i migrationsmønsteret for DNA  
heteroduplexer i non-denaturerende gel elektroforese, hvilket er blevet påvist  
for 3 bp (F508) mutationen og i andre eksperimentelle systemer [Nagamine et  
al, Am. J. Hum. Genet. 45: 337 - 339 (1989)]. Alternativt kunne en  
25 fremgangsmåde til påvisning af en mutation, der omfatter en enkelt base  
substitution eller anden lille ændring, kunne baseres på differential primer  
længde i en PCR. For eksempel kunne én ikke-varierende primer anvendes  
foruden en primer specifik for en mutation. PCR produkterne af de normale og  
mutante gener kan dernæst differentielt påvises i acrylamid geler.

30 Sekvensændringer ved specifikke lokaliseringer kan også afsløres ved hjælp  
af nuclease beskyttelse assays, såsom RNase (Myers, supra) og S1  
beskyttelse (Berk, A. J., og P. A. Sharpe Proc. Nat. Acad. Sci. U. S. A. 75:  
1274 (1978)], den kemiske spaltningsfremgangsmåde (Cotton, supra) eller  
35 den ligase-medieret påvisningsprocedure (Landegren, supra).



Foruden traditionel gel-elektroforese og blot-hybridisering fremgangsmåder kan DNA fragmenter også visualiseres ved hjælp af fremgangsmåder, hvor de individuelle DNA prøver ikke er immobiliseret på membraner. Proben og target-sekvenserne kan begge være i opløsning eller probesequensen kan  
5 være immobiliseret [Saiki et al, Proc. Natl. Acad. Sci USA, 86: 6230 - 6234(1989)]. En række påvisningsfremgangsmåder, såsom autoradiografi, der involverer radioisotoper, direkte påvisning af radioaktivt henfald (i tilstedeværelsen eller fraværet af scintillans), spektrofotometri, der involverer kolorigene reaktioner og fluorometri, der involverer fluorogene reaktioner, kan  
10 anvendes til at identificere specifikke individuelle genotyper.

Eftersom mere end én mutation er forventet i CF genet, er et multipelt system en ideal protokol til screening af CF bærere og påvisning af specifikke mutationer. For eksempel kan en PCR med multiple, specifikke oligonucleotid-primere og hybridiseringsprober anvendes til at identificere alle mulige  
15 mutationer på samme tid (Chamberlain et al. Nucleic Acids Research 16: 1141 - 1155 (1988)). Proceduren kan involvere immobiliserede sekvens-specifikke oligonucleotider prober (Saiki et al, supra).

## 20 5.2 PÅVISNING AF DEN STØRRE MUTATION

Disse påvisningsfremgangsmåder kan anvendes til prænatal diagnose under anvendelse af amnionvæske celler, chorion-villi biopsi eller sortering af føtale celler fra maternel cirkulation. Testen for CF bærere i populationen kan  
25 inkorporeres som en essentiel komponent i et storskala genetisk testningsprogram for almindelige sygdomme.

Ifølge en udførelsesform ifølge den foreliggende opfindelse kan den portion af det DNA segment, der er informativt for en mutation, såsom mutationen ifølge  
30 denne udførelsesform, det vil sige, den portion, der umiddelbart omgiver F508 deletionen, dernæst amplificeres ved at anvende standard PCR teknikker [som anført i oversigt i Landegren, Ulf, Robert Kaiser, C. Thomas Caskey, og Leroy Hood, DNA Diagnostics - Molecular Techniques og Automation, i Science 242: 229 - 237 (1988)]. Det overvejes, at den portion af det DNA segment, der  
35 anvendes, kan være et enkelt DNA segment eller en blanding af forskellige DNA segmenter. En detaljeret beskrivelse af denne teknik følger nu.

En specifik region af genomisk DNA fra den person eller det foster, der skal screenes. En sådan specifik region er defineret ved hjælp af oligonucleotidet primere C16B (5'GTTTTCTGGATTATGCCTGGGCAC3') og C16D (5'GTTGGCATGCTTTGATGACGCTTC3').

5

De specifikke regioner blev amplificeret ved hjælp af polymerase kæde reaktionen (PCR). 200 - 400 ng af genomisk DNA, fra enten dyrkede lymfoblaster eller perifere blod prøver fra CF individer og deres forældre, blev anvendt i hver PCR med de nedenfor viste oligonucleotid primere.

10 Oligonucleotiderne blev oprenset med Oligonucleotide Purification Cartridges™ (Applied Biosystems) eller NENSORB™ PREP søjler (Dupont) med procedurer anbefalet af leverandørerne. Primerne blev annealet ved 62 °C i 45 sekunder, ekstenderet ved 72 °C i 120 sekunder (med 2 units af Taq DNA polymerase) og denatureret ved 94 °C i 60 sekunder, i 28 cyklusser

15 med en final cyklus af 7 min for ekstension i en Perkin-Elmer/Cetus automatisk thermocycler med et Step-Cycle program (transition setting ved 1,5 min). Portioner af PCR produkterne blev adskilt ved hjælp af elektroforese på 1,4 % agarosegeler, overført til Zetabind; (Biorad) membran ifølge standard procedurer. De to oligonucleotidprober i figur 15 (10 ng hver) blev mærket

20 separat med 10 units af T4 polynucleotid kinase (Pharmacia) i en 10 µl reaktion, der indeholder 50 mM Tris-HCl (pH 7,6), 10 mM MgCl<sub>2</sub>, 0,5 mM dithiothreitol, 10 mM spermidine, 1 mM EDTA og 30-40 µCi af γ[<sup>32</sup>P] - ATP i 20 - 30 min ved 37 °C. De uinkorporerede radionucleotider blev fjernet med en Sephadex G-25 søjle før brug. Hybridiseringsbetingelserne var som beskrevet tidligere (J. M. Rommens et al Am. J. Hum. Genet. 43, 645 (1988)), bortset fra

25 at temperaturen var 37 °C. Membranerne blev vasket to gange ved stuetemperatur med 5 x SSC og to gange ved 39 °C med 2 x SSC (1 x SSC = 150 mM NaCl og 15 mM Na citrat). Autoradiografi blev udført ved stuetemperatur natten over. Autoradiografer viser hybridiseringsresultaterne

30 for genomisk DNA med de 2 specifikke oligonucleotidprober som vist i figur 15. Probe C påviser den normale DNA sekvens og Probe F påviser den mutante sekvens. Genomisk DNA prøve fra hvert familiemedlem blev amplificeret ved hjælp af polymerase kæde reaktionen, og produkterne blev adskilt ved hjælp

35 af elektroforese på en 1,4 % agarose gel og dernæst overført til Zetabind (Biorad) membran ifølge standard procedurer. Vand-blank og plasmid DNA, T16 og C16, svarende til henholdsvis den normale sekvens (N) og F508 deletionen (CF), blev indbefattet som kontroller.

3 bp deletionen blev også afsløret ved hjælp af polyacrylamidgel elektroforese. Når PCR'en frembragt ved hjælp af den oven for nævnte C16B og C16D primere blev anvendt til en 8 % polyacrylamidgel, underkastet elektroforese i 2 timer ved 20 V/cm i en 90 mM Tris-borat-buffer (pH 8,3), var DNA fragmenter med en anderledes mobilitet klart påviselig for individer uden 3 bp deletionen heterozygote eller homozygote for deletionen. Derudover blev et ekstra DNA bånd, formodentlig heteroduplex'en mellem normale og mutante DNA strenge, bemærket i heterozygote. Lignende ændring i gel mobilitet for heteroduplexer dannet under PCR er også blevet rapporteret for eksperimentelle systemer, hvor små deletioner er involveret (Nagamine et al supra). Disse mobilitet skift kan anvendes som basis for de non-radioaktive genetiske screeningstests.

### 5.3 CF SCREENINGSPROGRAMMER

Det anerkendes, at kun 70 % af bærerne kan påvises under anvendelse af de specifikke F508 prober ifølge denne særlige udførelsesform ifølge den foreliggende opfindelse. Hvis således et testet individ er ikke en bærer under anvendelse af F508 proberne, kan deres bærer status ikke udelukkes, de kan bær en anden mutation som tidligere bemærket. Hvis imidlertid både individet og ægtefællen til det testede individ er en bærer for F508 mutationen, kan det fastslås med vished, at de er et par i risiko. Sekvensen for genet som beskrevet heri er en essentiel forudsætning for bestemmelsen af de andre mutationer.

Prænatal diagnose er en logisk udvidelse af bærer-screening. Et par kan identificeres som at være i risiko for at få barn med cystisk fibrose på én af to måder: hvis de allerede har et barn med cystisk fibrose, er de begge, per definition, obligate bærere af sygdommen, og hvert efterfølgende barn har en 25 % risiko for at blive ramt af cystisk fibrose. En større fordel ifølge den foreliggende opfindelse eliminerer behovet for familiestamtræ-analyse, mens, ifølge den foreliggende opfindelse, et gen-mutation screening program som beskrevet ovenfor eller anden tilsvarende fremgangsmåde kan anvendes til at identificere en genetisk mutation, der fører til et protein med ændret funktion.

Dette er ikke afhængigt af forudgående fastsættelse af familien via et ramt barn. Føtale DNA prøver kan for eksempel opnås, som tidligere nævnt, fra

amnionvæske celler og chorion-villi prøver. Amplificering ved hjælp af standard PCR teknikker kan dernæst udføres på denne template DNA.

5 Hvis begge forældre er vist at være bærere med F508 deletionen, ville fortolkningen af resultaterne være det følgende. Hvis der er hybridisering af den føtale DNA til den normale (ingen deletion, som vist i figur 15) probe, vil fosteret ikke være ramt med cystisk fibrose, selv om det kan være en CF bærer (50 % sandsynlighed for hver foster af et par i risiko). Hvis den føtale DNA kun hybridiserer til F508 deletions-proben og ikke til den normale probe  
10 (som vist i figur 15), vil fosteret være ramt af cystisk fibrose.

Det anerkendes, at for denne og andre mutationer i CF genet kan en række af forskellige specifikke procedurer anvendes til at tilvejebringe en komplet diagnose for alle potentielle CF bærere eller patienter. En komplet beskrivelse  
15 af disse procedurer er beskrevet senere.

Den foreliggende opfindelse tilvejebringer derfor en fremgangsmåde og et kit til bestemmelse af, om et individ er en CF bærer eller CF patient. I sammendrag omfatter screeningsfremgangsmåden trinene med:  
20

at tilvejebringe en biologisk prøve fra det individ, der skal screenes; og tilvejebringelse af et assay til påvisning i den biologiske prøve, af tilstedeværelsen af mindst et medlem fra gruppen bestående af det normale CF gen, normal CF genprodukter, et mutant CF gen, mutante CF  
25 genprodukter og blandinger deraf.

Fremgangsmåden kan yderligere kendetegnes ved at indbefatte mindst én yderligere nucleotidprobe, der er et anderledes DNA sekvens fragment af for eksempel DNA'en i figur 1, eller et anderledes DNA sekvens fragment af  
30 humant kromosom 7 og lokaliseret til hver side af DNA sekvensen i figur 1.

Et kit, ifølge en udførelsesform ifølge den foreliggende opfindelse, egnet til anvendelse i screeningsteknikken og til undersøgelse for tilstedeværelsen af CF genet ved hjælp af et immunoassay omfatter:  
35

(a) et antistof, der specifikt binder til et genprodukt af CF genet;

(b) reagensmidler til påvisning af bindingen af antistoffet til genproduktet; og

(c) idet antistoffet og reagensmidlerne hver er til stede i mængder, der er effektive til at udføre immunoassayet.

5

Kittet til undersøgelse for tilstedeværelsen af CF genet kan også være tilvejebragt ved hjælp af hybridiseringsteknikker.

Kittet omfatter:

10

(a) en oligonucleotideprobe, der specifikt binder til CF genet;

(b) reagensmidler til påvisning af hybridiseringen af oligonucleotideproben til CF genet; og

15

(c) idet proben og reagensmidlerne hver er til stede present i mængder, der er effektive til at udføre hybridiseringsassayet.

#### 5.4 ANTISTOFFER TIL AT PÅVISE CFTR

20

Som nævnt er antistoffer mod epitoper inden for CFTR proteinet rejst for at tilvejebringe omfattende information om karakteristikaene for proteinet og anden værdifuld information, der indbefatter:

25

1. At muliggøre visualisering af proteinet i celler og væv, hvor det udtrykkes ved hjælp af immunoblotting ("Western blots") efter polyacrylamidgel elektroforese. Dette tillader en estimation af den molekylære størrelse for det modne protein, der indbefatter bidraget fra cellerne i post-translationelt tilsatte dele, herunder for eksempel oligosaccharidkæder og fosfatgrupper. Immunocytokemiske teknikker, der indbefatter immunfluorescens og immunelektronmikroskopi, kan anvendes til at etablere den subcellulære lokalisering af proteinet i cellemembraner. Antistofferne kan også anvendes til at tilvejebringe en anden teknik til påvisning af en hvilken som helst af de øvrige CF mutationer, som resulterer i syntesen af et protein med en ændret størrelse.

35

2. Antistoffer mod forskellige domæner i proteinet kan anvendes til at bestemme det topologiske arrangement af proteinet i cellemembranen. Dette tilvejebringer information om segmenter af det protein, som er tilgængeligt for eksternt tilsatte modulerende midler til formål inden for lægemiddelterapi.

5

3. Struktur-funktions-forholdet af portioner af proteinet kan undersøges ved at anvende specifikke antistoffer. For eksempel er det muligt at indføre i celler antistoffer, der genkender hver af de ladede cytoplasmatiske loops, der forbinder de transmembrane sekvenser såvel som portioner af de nucleotid-bindende folder og R-domænet. Disse antistoffers indflydelse på funktionelle parametre for proteinet tilvejebringer indsigt i celle-regulatoriske mekanismer og foreslår potentielt midler til modulering af aktiviteten af det defekte protein hos en CF patient.

10

15

4. Antistoffer med den passende aviditet muliggør også immunopræcipitering og immunoaffinitetsoprensning af proteinet. Immunopræcipitering vil fremme karakterisering af syntese og post-translational modifikation, herunder ATP binding og phosphorylering. Oprensning vil være påkrævet for studier af proteinstruktur og for rekonstitution af dets funktion, såvel som protein-baseret terapi.

20

Med henblik på at fremstille antistofferne, er fusionsproteiner, der indeholder definerede portioner af CFTR polypeptider, blevet syntetiseret i bakterier ved hjælp af ekspression af korresponderende DNA sekvenser i en egnet kloningsvehikel, mens mindre peptider blev syntetiseret kemisk som beskrevet i tabel 8. Fusionsproteinerne blev oprenset, for eksempel ved hjælp af affinitetskromatografi på glutathion-agarose, og peptiderne blev koblet til et bærerprotein (hemocyanin), blandet med Friends adjuvans og injiceret i kaniner. Efter booster-injektioner med intervaller på 2 uger, blev kaninerne tappet, og sera blev isoleret. De farvede fusionsproteiner er vist i figur 19a. Bane 1, uinduceret kontrol plasmid; bane 2, IPTG-induceret kontrol plasmid, der udtrykker netop glutathion-S-transferase (GST); bane 3, affinitetsoprenset GST bånd ved 27 kilodalton (kD); bane 4 er uinduceret, bane 5 er induceret, og bane 6 er det oprensede fusionsprotein #1 i tabel 8. I figur 19b er gel-elektroforesen af lysater fra bakterier transformeret med pGEX plasmider, der indeholder fusionsproteiner #5 i tabel 8 for banerne 1 og 2 og fusionsproteinerne #2 i tabel 8 for banerne 3 og 4. Bane 1 i figur 19b er for det

25

30

35

uinducerede plasmid, mens bane 2 er for det inducerede plasmid til  
ekspression af fusionsproteinet #5. Bane 3 i figur 19b er for det uinducerede  
plasmid, mens bane 4 er for det inducerede plasmid til ekspression af  
fusionsproteinet #2. Immunoblots af fusionsprotein #1 probet med antisera  
5 opnået fra anden tapning af to forskellige kaniner er vist i figur 20. Farvningen  
er med alkalisk-phosphatase konjugeret sekundært antistof [Blake et al, Anal.  
Biochem. 136:175, (1984)]. Begge af disse immunsera farver 32 kD  
fusionsproteinet, mens de præimmune sera ikke gør. Figur 21 viser  
reaktiviteten af ét af disse immunsera med et bånd på ca. 200 kD i størrelse i  
10 membraner isoleret fra T-84 coloncarcinom celler, der udtrykker CFTR  
transcript'et på et højt niveau. Dette bånd er i det størrelsesområde, der vil  
forventes for det CFTR protein, der har en forudsagt molekylvægt på 169 kD  
forud for post-translationelle modifikationer.

15 Sera fra kaniner immuniseret med LKH konjugatet af peptid #2 blev screenet  
igen både rent peptid og KLH som vist i figur 22. I denne figur betegner H  
hemocyanin; P1, peptid #1; P2, peptid #2. Mængder af protein eller peptid  
prykket i ng er vist. Dette antiserum påviser så lidt som 1 ng af peptidet og  
reagerer overhovedet ikke med kontrol peptid #1.

20 Det er således muligt at rejse polyklonale antistoffer specifikke for både  
fusionsproteiner, der indeholder portioner af CFTR proteinet, og CFTR  
peptiderne svarende til korte segmenter af dets sekvens. Tilsvarende kan mus  
injiceres med KLH konjugater af peptiderne 1, 2 og 7 fra tabel 8 med henblik  
25 på at initiere produktionen af monoklonale antistoffer mod disse segmenter af  
CFTR protein. Monoklonale antistoffer kan være tilsvarende rejst mod andre  
domæner af CFTR proteinet.

30 Som for frembringelsen af polyklonale antistoffer er immunogener for  
rejsningen af monoklonale antistoffer (mAb'er) mod CFTR proteinet bakterielle  
fusionsproteiner (Smith et al, Gene 67:31 (1988)], der indeholder portioner af  
CFTR polypeptidet eller syntetiske peptider svarende til korte (12 til 25  
aminosyrer i længde) segmenter af sekvensen. Den essentielle metodologi er  
den af Kohler og Milstein [Nature 256: 495 (1975)].

35 Balb/c mus immuniseres ved hjælp af intraperitoneal injektion med 500 µg af  
rent fusionsprotein eller syntetisk peptid i inkomplet Freund's adjuvans. En

anden injektion er givet efter 14 dage, en tredje efter 21 dage og en fjerde efter 28 dage. Individuelle dyr, der er immuniseret således, aflives én, to og fire uger efter den endelige injektion. Milte fjernes, deres celler dissocieres, opsamles og fusioneres med Sp2/O-Ag14 myelomceller ifølge Gefer et al, Somatic Cell Genetics 3: 231 (1977). Fusionsblandingen fordeles i dyrkningsmedium selektivt for opformeringen af fusionerede celler, der er dyrket, indtil de er omkring 259 konfluerende. På dette tidspunkt testes kultursupernatanter for tilstedeværelsen af antistoffer, der reagerer med et særligt CFTR antigen. Et alkalisk phosphatase mærket anti-muse sekundært antistof anvendes dernæst til påvisning af positive. Celler fra positive kulturbrønde ekspanderes dernæst i kultur, deres supernatanter opsamles til yderligere testning, og cellerne opbevares dybfrossen i kryoprotectant-holdigt medium. For at opnå store mængder af et mAb injiceres producent-celler i peritoneum ved  $5 \times 10^6$  celler per dyr, og der opnås ascitesvæske. Oprensning er ved hjælp af kromatografi på Protein G- eller Protein A-agarose ifølge Ey et al, Immunochemistry 15: 429 (1977).

Reaktivitet af disse mAb'er med CFTR proteinet bekræftes ved hjælp af polyacrylamidgel elektroforese af membraner isoleret fra epiteliale celler, hvor det udtrykkes, og immunoblotting [Towbin et al, Proc. Natl. Acad. Sci. USA 76: 4350 (1979)].

Foruden anvendelsen af monoklonale antistoffer specifikke for hver af de forskellige domæner i CFTR proteinet til at probe deres individuelle funktioner, anvendes andre mAb'er, der kan skelne mellem de normale og de mutante former af CFTR protein, til at påvise det mutante protein i epiteliale celleprøver opnået fra patienter, såsom nasal mucosa biopsi "børstninger" [R. DeLoughand J. Rutland, J. Clin. Pathol. 42, 613 (1989)] eller hudbiopsi prøver, der indeholder svedkirtler.

Antistoffer, der er i stand til denne skelnen, opnås ved hjælp af differentielt screenende hybridomer fra parrede sæt af mus immuniseret med et peptid, der indeholder phenylalaninen ved aminosyreposition 508 (e.g. GTIKENIIFGVSY) eller et peptid, der er identisk bortset fra fraværet af F508 (GTIKENIIGVSY). mAb'er, der er i stand til at genkende de andre mutante former af CFTR protein til stede i patienter foruden eller i stedet for F508



deletion er opnået under anvendelse af lignende strategier for produktion af monoklonalt antistof.

- 5 Antistoffer mod normale og CF versioner af CFTR protein og af segmenter deraf anvendes i diagnostisk immunocytokemisk og immunofluorescens lysmikroskopi og immunoelektronmikroskopi med henblik på at demonstrere vævsfordelingen, den cellulære og subcellulære fordeling af CFTR inden for organerne hos CF patienter, bærere og non-CF individer.
- 10 Antistoffer anvendes til at terapeutisk modulere ved at fremme aktiviteten af CFTR proteinet hos CF patienter og i celler fra CF patienter. Mulige måder til en sådan modulation vil kunne involvere stimulation på grund af krydsbinding af CFTR protein molekyler med multivalente antistoffer i analogi med stimulation af visse celleoverflade membran-receptorer, såsom insulin
- 15 receptoren [O'Brien et al, Euro. Mol. Biol. Organ. J. 6: 4003 (1987)], epidermal vækstfaktor receptor [Schreiber et al, J. Biol. Chem. 258: 846 (1983)] og T-celle receptorassocierede molekyler, såsom CD4 [Veillette et al Nature, 338: 257 (1989)].
- 20 Antistoffer anvendes til at styre tildelingen af terapeutiske midler til de celler, der udtrykker defekt CFTR protein i CF. Til dette formål inkorporeres antistofferne i en vehikel, såsom et liposom [Matthay et al, Cancer Res. 46: 4904 (1986)], der bærer det terapeutiske middel, såsom et lægemiddel eller det normale gen.

25

TABEL 8  
CFTR FRAGMENTER ANVENDT TIL AT REJSE ANTISTOFFER

GST <sup>a</sup> fusionsproteiner, der indeholder CFTR rester	CFTR domæne fra Fig. 13
1. 204-249	TM3, Ext. 2, TMA
2. 347-698	NBF-1, N-term 1/2 R-domæne
3. 710-757	Neg. ladet midte af R-domæne
4. 758-796	Pos. ladet segment af R-domæne
5. 1188-1480	C-term. cyto. domæne med NBF-2

5 KLH<sup>b</sup> konjugater, der indeholder CFTR peptider:

1. 28-45	N-term. cytoplasmatisk
2. 58-75	N-term. cytoplasmatisk
3. 104-117	Første ekstracellulære
4. 139-153	Anden cytoplasmatisk
5. 279-294	N-term. af tredje cytoplasmatisk
6. 500-512	NBF-1; omkring F508 deletionen
7. 725-739	Ladet midte af R-domæne
8. 933-946	Femte cytoplasmatisk
9. 1066-1084	sjette cytoplasmatisk

a. restriktionsfragmenter, der koder for disse fragmenter ligeret til 3' ende af glutathion S-transferase (GST) af *Schistosoma japonicum* i pGEX plasmid ekspressionsvektor som identificeret i Smith et al, Gene 67: 31, (1988).

10

b. Peptider koblet via en N-terminal cystein til bærer-proteinet keyhole limpet hemocyanin (RLH) ifølge Green et al Cell 28: 477 (1982). TM betegner transmembrane sekvenser.

## 15 5.5 RFLP ANALYSE

Den foreliggende opfindelse tilvejebringer en række af fordele, der stammer direkte fra opdagelsen og karakteriseringen af det CF gen, der har umiddelbar praktisk anvendelse. Aminosyresekvensen for CFTR tilvejebringer indsigt i proteinets struktur og funktion såvel som de molekylære mekanismer, hvor

20

CFTR deltager, og som er defekt ved cystisk fibrose. Denne information muliggør frembringelsen af yderligere redskaber og koncepter, i forskning om og terapi for denne sygdom.

5 Bærerpåvisning, DNA diagnose og familierådgivning er nogle af  
anvendelserne ifølge den foreliggende opfindelse. Tidligere DNA-baseret  
genetisk testning for CF har primært været tilgængelig for familier med ramte  
børn og to deres nære slægtninge. Viden om CF mutationerne på DNA  
10 sekvens niveauet tillader testning af et hvilket som helst tilfældigt individ; vores  
estimat viser, at 46 % af CF patienter uden en tidligere familiehistorie kan  
diagnosticere nøjagtigt ved hjælp af DNA analyse, og 68 % af CF bærerne i  
populationen kan identificeres via F508 deletionen.

15 Givet at bærer-frekvensen i den Nordamerikanske population er ca. 1 ud af 20,  
kan det lade sig gøre at screene alle kvinder og/eller mænd i den fødedygtige  
alder, for eksempel for deres bærer-status. Bærer-påvisning under anvendelse  
af prober specifikke for F508 deletionen vil udpege 70 % af bærerne. De  
resterende bærere vil påvises ved hjælp af et batteri af prober specifikke for de  
forskellige haplotype-grupper identificeret ovenfor.

20 Eftersom F508 deletionen udgør omtrentligt 70 % af alle CF mutationer, kan  
RFLP analyse anvendes som supplement til den direkte deletionstestning for  
familiemedlemmer eller nære slægtninge til CF patienter. Omkring 55 % af CF  
de forældre, der ikke bærer F508 mutation forventes at være informative for  
25 DNA markøren JG2E1 (KM19) [Kerem et al Am. J. Hum. Genet 44:827 - 834  
(1989); Estivill et al, Genomics 1: 257 (1987)] baseret på retrospektiv analyse  
af vore CF binding familier; en yderligere 39 % ville være informative, hvis E6  
(Taq I) (Kerem et al, supra og J3.11 (Msp I) [Wainright et al Nature (1985)]  
også blev testet; praktisk talt alle forældre ville være informative, hvis H2.3  
30 (XV2C-Taq I) [Kerem et al, supra; Estivill et al, Nature (1987)], E2.6 (E.9) (Msp  
I) [probe tilgængelig efter anmodning), E4.1 (Mp6d.9) (Msp I) [probe  
tilgængelig efter anmodning; Estivill et al, Am. J. Hum. Genet. (1989)], J44  
(E3.1) (Xba I) [probe tilgængelig efter anmodning] og metD (Ban I) [Spence et  
al, Am. J. Hum. Genet (1986), [ATCC #40219] blev indbefattet.

35 anvendeligheden af disse prober ligger i det faktum, at de genkender  
polymorfe restriktionssites. Således er proberne typisk ikke defineret ved deres

sekvens tværs gennem det særlige polymorfe site, men kan snarere benyttes baseret på viden om flankerende sekvenser, der tillader polymerase kæde reaktion (PCR) frembringelse af den region, der er tale om, hvilket ville være kendt af fagmanden inden for det tekniske område.

5

For eksempel er proben E2.6 (Msp I) komplet defineret af to flankerende oligomerer:

5'GTGATCCAGTTTGCTCTCCA3' og 5'GGAATCACTCTTCCTGATAT3'.

10

Anvendelse af denne E2.6 PCR frembragt3 probe til påvisning af en Msp I polymorfisme vil påvise to forskellige alleler: enten ét 850 bp fragment eller et 490 bp og et 360 bp fragment, afhængigt af tilstedeværelsen eller fraværet af Msp I site't. Tilsvarende er proben J44 (E3.1) (Xba I) komplet defineret af to flankerende oligomerer:

15

5'CAATGTGATTGGTGAAACTA3' og 5'CTTCTCCTCCTAGACACCTGCAT3'.

Anvendelse af denne J44 (E3.1) PCR frembragte probe til påvisning af en Xba I polymorfisme vil påvise to forskellige alleler: enten et 860 bp fragment eller et 610 bp og et 250 bp fragment, afhængigt af tilstedeværelsen eller fraværet af Xba I site't.

20

De forbundne RFLP'er kan også anvendes til risikoberegning for individer, som ikke bærer F508 deletionen. En generel procedure for risikoestimat er blevet drøftet i Beaudet et al Am. J. Hum. Genet 44: 319 - 326].

25

Til prænatal diagnose kan mikrovillær intestinal enzym analyse [Brock, Lancet 2: 941 (1983)] udføres til at forøge konfidens for diagnose i tilfælde, hvor DNA diagnose er inkonklusiv.

30

DNA diagnose anvendes for tiden til at fastsætte, om et foster vil blive født med cystisk fibrose, men historisk er dette kun blevet udført, efter at et givet forældrepar allerede havde fået ét barn med cystisk fibrose, hvilket identificerer dem som obligate bærere. I kombination med bærer-påvisning som beskrevet ovenfor vil imidlertid DNA diagnose for alle graviditeter hos bærer-par være muligt. Hvis forældrene allerede har fået et barn med cystisk

35

5 fibrose, kan en ekstenderet haplotype analyse udføres på fosteret, og således vil procentdelen af falsk positive eller falsk negative være stærkt reduceret. Hvis forældrene ikke allerede har fået et ramt barn, og DNA diagnosen på fosteret udføres på basis af bærer-påvisning, kan haplotype-analyse stadig udføres.

10 Selv om det i mange år er blevet antaget, at der er en hel del klinisk heterogenitet ved cystisk fibrose sygdommen, kommer det nu frem, at der er to generelle kategorier, betegnet pancreatisk sufficiens (CF-PS) og pancreatisk insufficiens (CF-PI). Hvis de mutationer, der er relateret til disse sygdomskategorier er velkarakteriserede, kan man associere en særlig mutation med en klinisk fænotype af sygdommen. Dette tillader ændringer i behandlingen af hver patient. Således vil mutationens beskaffenhed i en vis grad forudsige prognosen hos patienten og pege på en specifik behandling.

15

#### 6.0 MOLEKYLÆR BIOLOGI FOR CYSTISK FIBROSE

20 Det postulat, at CFTR kan regulere aktiviteten af ion-kanaler, især den udadgående korrigerende C1 kanal impliceret som den funktionelle defekt i CF, kan testes ved injektionen og translationen af fuld længde *in vitro* transkriberet CFTR mRNA i *Xenopus* oocyter. De deraf følgende ændringer i ion-strømme tværs gennem oocyt-membranen kan måles, eftersom potentialet er clampet ved en fikseret værdi. CFTR kan regulere endogene oocyt-kanaler, eller det kan være nødvendige også at indføre epitelial celle RNA til at styre translationen af kanal-proteiner. Anvendelse af mRNA, der koder for normal og for mutant CFTR som tilvejebragt ifølge den foreliggende opfindelse, gør disse eksperimenter mulige.

25

30 Andre måder for ekspression i heterolog celle system fremmer også dissektion af struktur-funktion forhold. Den komplette CFTR DNA sekvens ligret ind i en plasmid ekspressionsvektor anvendes til at transficere celler, således at dens indflydelse på ion-transport kan fastsættes. Plasmid ekspressionsvektorer, der indeholder en del af den normale CFTR sekvens langs med portioner af modificeret sekvens ved selekterede sites kan anvendes i *in vitro* mutagenese-eksperimenter udført med henblik på at identificere de portioner af CFTR proteinet, der er afgørende for regulatorisk funktion.

35

## 6.1 EKSPRESSION AF DNA SEKVENNS

DNA sekvensen kan manipuleres i studier med henblik på at forstå  
ekspressionen af genet og dets produkt, og med henblik på at opnå produktion  
5 af store mængder af proteinet for funktionel analyse, antistofproduktion og  
patientterapi. Ændringerne i sekvensen kan eller kan ikke ændre  
ekspressionsmønsteret i form af relative mængder, vævsspecificitet og  
funktionelle egenskaber. De partielle eller fuld-længde cDNA sekvenser, der  
10 koder for det foreliggende protein, umodificeret eller modificeret, kan være  
ligeret til bakterielle ekspressionsvektorer, såsom pRIT [Nilsson et al. EMBO J.  
4: 1075 - 1080 (1985)], pGEX [Smith og Johnson, Gene 67: 31 - 40 (1988)]  
eller pATH (Spindler et al. J. Virol. 49: 132 - 141 (1984)] plasmider, der kan  
være indført i E. coli celler til produktion af de korresponderende proteiner,  
15 som kan være isoleret ifølge de tidligere drøftede protein-  
oprensningsprocedurer. DNA sekvensen kan også være overført fra dens  
eksisterende sammenhæng med andre kloningsvehikler, såsom andre  
plasmider, bakteriofager, cosmider, animalske virus, gær kunstige  
kromosomer (YAC) [Burke et al. Science 236: 806 - 812, (1987)], somatiske  
20 celler og andre simple eller komplekse organismer, såsom bakterier, svampe  
(Timberlake og Marshall, Science 244: 1313 - 1317 (1989), invertebrate,  
planter (Gasser og Fraley, Science 244: 1293 (1989), og svin [Pursel et al.  
Science 244: 1281 - 1288 (1989)].

Til ekspression i mammale celler kan cDNA sekvensen være ligeret til  
25 heterologe promotorer, såsom simian virus (SV) 40, promotor i pSv2 vektoren  
[Mulligan og Berg, Proc. Natl. Acad. Sci USA, 78: 2072 - 2076 (1981)] og  
indført i celler, såsom abe COS-1 celler, [Gluzman, Cell, 23: 175 - 182 (1981)],  
med henblik på at opnå forbigående eller langtids ekspression. Den stabile  
integration af den kimæriske gen-konstruktion kan opretholdes i mammale  
30 celler ved biokemisk selektion, såsom neomycin (Southern og Berg, J. Mol.  
Apply Genet. 1: 327 - 341 (1982)] og mycophenolsyre [Hulligan og Berg,  
supra].

DNA sekvenser kan manipuleres med standard procedurer, såsom  
35 restriktionsenzym-spaltning, udfyldning med DNA polymerase, deletion ved  
hjælp af exonuclease, ekstension ved hjælp af terminal deoxynucleotid-  
transferase, ligering af syntetiske eller klonede DNA sekvenser, site-styret

sekvens-ændring via enkeltstrenget bakteriofag intermediaær eller med anvendelsen af specifikke oligonucleotider i kombination med PCR.

5 cDNA sekvensen (eller portioner afledt derfra), eller et mini gen (en cDNA med en intron og dens egen promotor) indføres i eukaryote ekspressionsvektorer ved hjælp af traditionelle teknikker. Disse vektorer designes til at tillade transcriptionen af cDNA'en i eukaryote celler ved at tilvejebringe regulatoriske sekvenser, der initierer og forøger transcriptionen af cDNA'en og sikrer dens korrekte splejsning og polyadenylering. Vektorer, der indeholder promotor- og enhancer-regionerne af simian virus (SV)40 eller lang terminal gentagelse (repeat) (LTR) af Rous Sarcom virus og polyadenylerings- og splejsningssignal fra SV 40 er let tilgængelige (Mulligan et al Proc. Natl. Acad. Sci. USA 78: 1078 - 2076, (1981); Gorman et al Proc Natl. Acad. Sci USA 79: 6777 - 6781 (1982)). Alternativt kan den CFTR endogene promotor anvendes. Niveauet for 15 ekspression af cDNA'en kan manipuleres med denne type af vektor, enten ved at anvende promotorer, der har forskellige aktiviteter (for eksempel kan baculovirus pAC373 udtrykke cDNA'er på høje niveauer i S. frungiperda celler [M; D. Summers og G. E. Smith i, Genetically Altered Viruses and the Environment (B. Fields, et al, eds.) vol. 22 no 319 - 328, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, 1985) eller ved at anvende 20 vektorer, der indeholder promotorer, som kan underkastes modulation, for eksempel den glucocorticoid-responsive promotor fra muse mammatumour virus'en [Lee et al, Nature 294: 228 (1982)]. Ekspressionen af cDNA'en kan monitoreres i recipient-cellerne 24 til 72 timer efter indføring (forbigående 25 ekspression).

Derudover indeholder nogle vektorer selekterbare markører [såsom gpt [Mulligan et Berg supra] eller neo [Southern og Berg J. Mol. Appln. Genet 1: 327 - 341 (1982)] bakterielle gener, der tillader isolation af celler, ved hjælp af kemisk selektion, som har have stabile, langtids ekspression af vektorerne (og derfor cDNA'en) i recipientcellen. Vektorerne kan opretholdes i cellerne som episomale, frit replicerende enheder ved at anvende regulatoriske elementer af virusser, såsom papillom (Sarver et al Mol. Cell Biol. 1: 486 (1981)) eller Epstein Barr (Sugden et al Mol. Cell Biol. 5: 410 (1985)). Alternativt kan man 30 også producere cellelinjer, der har integreret vektoren i genomisk DNA. Begge disse typer af cellelinjer producerer genproduktet på en kontinuerlig basis. Man kan også producere cellelinjer, der har amplificeret antallet af kopier af

vektoren (og derfor også af cDNA'en) med henblik på at frembringe cellelinjer, der kan producere høje niveauer af genproduktet [Alt et al. J. Biol. Chem. 253: 1357 (1978)].

5 Overførslen af DNA til eukaryote, især humane eller andre mammale celler, er nu en traditionel teknik. Vektorerne indføres i recipient-cellerne som ren DNA (transfektion) ved for eksempel præcipitering med calciumphosphat [Graham og vander Eb, Virology 52: 466 (1973) eller strontiumphosphat [Brash et al Mol. Cell Biol. 7: 2013 (1987)], elektroporation [Neumann et al Embo J 1: 841  
10 (1982)], lipofektion [Felgner et al Proc Natl. Acad. Sci USA 84: 7413 (1987)], DEAE dextran [McCuthan et al, J. Natl Cancer Inst. 41: 351 1968)], mikroinjektion [Mueller et al Cell 15: 579 1978)], protoplast-fusion [Schafner, Proc Natl. Aca. Sci USA 72: 2163] eller pellet-bøsser [Klein et al, Nature 327: 70 (1987)]. Alternativt kan cDNA'en kan indføres ved hjælp af infektion med  
15 virus-vektorer. Der er udviklet systemer, der for eksempel benytter retrovirus, [Bernstein et al. Genetic Engineering 7: 235, (1985)], adenoviruser, [Ahmad et al J. Virol 57: 267 (1986)] eller Herpes virus [Spaete et al Cell 30: 295 (1982)].

20 Disse eukaryote ekspressionssystemer kan anvendes til mange studier af CF genet og CFTR produktet. Disse indbefatter for eksempel: (1) bestemmelse af at genet er korrekt udtrykt, og at alle post-translational modifikationer, som er nødvendige for fuld biologisk aktivitet, er blevet korrekt fuldført (2) at identificere af regulatoriske elementer lokaliseret i 5' regionen af CF genet og deres rolle i vævsreguleringen eller den temporale regulering af ekspressionen  
25 af CF genet (3) produktion af store mængder af det normale protein til isolation og oprensning (4) at anvende celler, der udtrykker CFTR proteinet som et assay system for antistoffer frembragt mod CFTR proteinet eller et assaysystem til at teste effektiviteten af lægemidler, (5) studium af funktionen af det normale komplette protein, specifikke portioner af proteinet eller af  
30 naturligt forekommende eller kunstigt fremstillede mutante proteiner. Naturligt forekommende mutante proteiner eksisterer i patienter med CF, mens kunstigt producerede mutante protein kan designes ved hjælp af site-styret sekvensændringer. Disse senere studier kan probe funktionen af en hvilken som helst ønsket aminosyrerest i proteinet ved at mutere de nucleotider, der  
35 koder for den pågældende aminosyre.



Under anvendelse af de ovennævnte teknikker kan de ekspressionsvektorer, der indeholder CF gen sekvensen eller fragmenter deraf indføres i humane celler, mammae celler fra andre arter eller non-mammae celler, alt efter hvad der er ønsket. Valget af celle er bestemt af formålet med behandlingen. For eksempel kan man anvende abe COS celler [Gluzman, Cell 23: 175 (1981)], der producerer høje niveauer af SV40 T antigenet og tillade replikeringen af vektorer, der indeholder SV40 origin for replikation, kan anvendes til at vise, at vektoren kan udtrykke proteinproduktet, eftersom funktion ikke er påkrævet. Lignende behandling kunne udføres med Kinesisk hamster ovarie (CHO) eller muse NIH 3T3 fibroblaster eller med humane fibroblaster eller lymfoblaster.

Den rekombinante kloningsvektor ifølge til den foreliggende opfindelse, omfatter da den selekterede DNA af DNA sekvenserne ifølge den foreliggende opfindelse til ekspression i en egnet vært. DNA'en er operativt forbundet i vektoren til en ekspressionskontrolsekvens i det rekombinante DNA molekyle, således at normalt CFTR polypeptid kan udtrykkes. Ekspressionskontrolsekvensen kan være udvalgt fra gruppen bestående af sekvenser, der kontrollerer ekspressionen af gener af prokaryote eller eukaryote celler og deres virusser og kombinationer deraf. Ekspressionskontrolsekvensen kan udvælges specifikt udvalgt fra gruppen bestående af lac systemet, trp systemet, tac systemet, trc systemet, større operator og promotor regioner af fag lambda, kontrolregionen af fd coat protein, de tidlige og sene promotorer af SV40, promotorer afledt fra polyom, adenovirus, retrovirus, baculovirus og simian virus, promotoren for 3-phosphoglycerat-kinase, promotorerne af gær sur phosphatase, promotoren af gær alfa-mating faktorer og kombinationer deraf.

Den værtscelle, der kan være transficeret med vektoren ifølge den foreliggende opfindelse, kan udvælges fra gruppen bestående af E. coli, Pseudomona, Bacillus subtilis, Bacillus stearothermophilus eller andre bacilli; andre bakterier; gær; svampe; insekt; muse eller andet dyr; eller planteværter; eller humane vævsceller.

Det forstås, at for de mutante DNA sekvens lignende systemer benyttes til at udtrykke og producere det mutante produkt.

## 6.2 PROTEINFUNKTIONSOVERVEJELSER

Med henblik på at studere funktionen af CFTR proteinet foretrækkes det at anvende epiteliale celler som recipienter, eftersom korrekt funktionel ekspres-  
5 sion kan kræve tilstedeværelsen af andre pathways eller genprodukter, der kun udtrykkes i sådanne celler. Celler, der kan anvendes, indbefatter for eksempel humane epiteliale cellerlinjer, såsom T84 (ATCC #CRL 248) eller PANC-1 (ATCC# CLL 1469), eller den T43 immortaliserede CF nasal epitel cellerlinje [Jettan et al, Science (1989)] og primære [Yanhoskes et al. Ann. Rev. Resp. Dis. 132: 1281 (1985)] eller transformeret [Scholte et al. Exo. Cell. Res. 182: 559 (1989)] human nasal polyp eller luftvejsceller, pancreatiske celler [Harris og Coleman J. Cell. Sci. 87: 695 (1987)], eller svedkirtel-celler [Collie et al. In Vitro 21: 597 (1985)] afledt fra normale eller CF individer. CF cellerne kan anvendes til at teste for den funktionelle aktivitet af  
10 mutante CF gener. For tiden tilgængelige funktionelle assays tilgængelig indbefatter studiet af bevægelsen af anioner (Cl eller I) tværs gennem cellemembraner som en funktion af stimulation af celler ved hjælp af midler, der hæver intracellulære AMP niveauer og aktiverer chlorid-kanaler [Stutto et al. Proc. Nat. Acad. Sci. U. S. A. 82: 6677 (1985)]. Andre assays indbefatter  
15 målingen af ændringer i cellulære potentialer ved hjælp af patch clamping af hele celler eller af isolerede membraner [Frizzell et al. Science 233: 558 (1986), Welsch og Liedtke Nature 322: 467 (1986)] eller studiet af ion-fluxer i epiteliale ark af konfluerende celler [Widdicombe et al. Proc. Nat. Acad. Sci. 82: 6167 (1985)]. Alternativt kunne RNA fremstillet fra CF genet injiceres i  
20 Xenopus oocyter. Oocytten vil translaterer RNA til protein og muliggøre dets studium. Eftersom andre mere specifikke assays er udviklet, kan disse også anvendes til studiet af transficeret CFTR protein funktion.

Eksperimenter med "domæne-switching" mellem CFTR og det humane multilægemedel-resistens P-glycoprotein kan også udføres for yderligere at  
30 studere CFTR proteinet. I disse eksperimenter er plasmid ekspressionsvektorer konstrueret ved hjælp af rutine teknikker fra fragmenter af d CFTR sekvensen og fragmenterne af sekvensen af P-glycoprotein ligeret sammen ved hjælp af DNA ligase, således at et protein, der indeholder de  
35 respektive portioner af disse to proteiner, vil blive syntetiseret af en værtscelle transficeret med plasmidet. Det sidstnævnte approach har den fordel, at mange eksperimentelle parametre associeret med multilægemedel-resistens

kan måles. Det er således nu muligt at fastsætte evnen hos segmenter af CFTR til at øve indflydelse på disse parametre.

5 Disse studier af indflydelsen af CFTR på ion-transport vil tjene til at bringe området for epitelial transport ind på den molekylære arena. Dette er det første transport-relaterede molekyle fra epiteliale celler, for hvilket den komplette primære struktur er vist. Viden om CFTR kan anvendes til at bedre forstå, på et molekylært niveau, karakteristikaene for den epiteliale cellemembran i dette område. For eksempel kan de molekyler, der er i tættest nærhed af CFTR, bestemmes ved hjælp af krydsbindingseksperimenter. Den hypotese, at rollen for CFTR er at regulere ion-kanaler, ville forudsige, at disse kanaler nødvendigvis ville falde i den pågældende kategori. De store, høj-kvalitet cDNA biblioteker konstrueret til kloningen af CFTR cDNA'er vil også være anvendelige til den molekylære kloning af cDNA'er for polypeptider, der udgør 10 andre epiteliale ion-transport systemer, herunder andre kanaler såvel som co-, 15 mod- og aktiv-transport systemer.

### 6.3 TERAPIER

20 Det forstås, at det væsentlige sigte med de forskellige biokemiske studier, der anvender sammensætningerne ifølge den foreliggende opfindelse, er udviklingen af terapier til at omgå eller overvinde CF defekten, under anvendelse af begge de farmakologiske og "gen-terapi" approaches.

25 I det farmakologiske approach søges lægemidler, der omgår eller overvinder CF defekten. Initialt kan forbindelser testes i det væsentlige tilfældigt, og screeningssystemer er påkrævet til at skelne blandt mange kandidat-forbindelser. Den foreliggende opfindelse tilvejebringer værtscelle-systemer, der udtrykker flere af de mutante CF gener, der er særligt velegnede til 30 anvendelse som første niveau screeningssystemer. Fortrinsvis anvendes der i screeningsfremgangsmåden et celledyrkningssystem, som benytter mammale celler (mest foretrukket humane celler) transficeret med en ekspressionsvektor, der omfatter en DNA sekvens, som koder for CFTR protein, der indeholder en CF-frembringende mutation, for eksempel F508 deletionen. Kandidat-lægemidler testes ved at inkubere cellerne i 35 tilstedeværelsen af kandidat-lægemidlet og måling af de cellulære funktioner, der er afhængige af CFTR, især ved at måle ion-strømme, hvor det

transmembrane potentiale er clampet ved en fikseret værdi. For rumme det store antal af assays er mere hensigtsmæssige assays imidlertid baseret på for eksempel anvendelsen af ion-sensitive fluorescens-farvestoffer. For at påvise ændringer i Cl-ion koncentration er SPQ eller dens analoger anvendelige.

Alternativt kunne et celle-frit system anvendes. Oprensset CFTR kunne rekonstitueres i kunstige membraner, og lægemidler kunne screenes i et celle-frit assay [Al-Aqwatt, Science, (1989)].

På det sekundære niveau er dyre-testning påkrævet. Det er muligt at udvikle en model for CF ved at interferere med den normale ekspression af modstykket til CF genet i et dyr, såsom musen. "Knock-out" af dette gen ved indføring af en mutant form deraf i kimcellelinjen af dyr vil tilvejebringe en stamme af dyr med CF-lignende syndromer. Dette muliggør testning af lægemidler, der viste sig lovende på det første celle-baserede screeningsniveau.

Efterhånden som yderligere viden er opnået om beskaffenheden af proteinet og dets funktion, vil det være muligt at forudsige strukturer af proteiner eller andre forbindelser, der interagerer med CFTR proteinet. Det vil atter muliggøre, at der bliver gjort visse forudsigelser om potentielle lægemidler, der vil interagere med dette protein og have en vis virkning på behandlingen af patienterne. I sidste ende kan sådanne lægemidler designes og syntetiseres kemisk på basis af strukturer forudsagt at være påkrævet til at interagere med domæner i CFTR. Dette approach er gennemgået i oversigt i Capsey og Delvatte, Genetically Engineered Human Therapeutic Drugs Stockton Press, New York, 1988. Disse potentielle lægemidler skal også testes i screeningssystemet.

### 6.3.1 PROTEINERSTATNINGS-TERAPI

Behandling af CF kan udføres ved at erstatte det defekte protein med normalt protein, ved modulering af funktionen i det defekte protein eller ved modificering af yderligere trin i den pathway, hvor CFTR deltager med henblik på at korrigere det fysiologiske abnormalitet.

For at være i stand til at erstatte det defekt protein med den normale version, skal man have rimeligt store mængder af rent CFTR protein. Rent protein kan opnås som beskrevet tidligere fra dyrkede celledsystemer. Tildeling af proteinet til de berørte luftvejsvæv vil kræve dets pakning i lipid-holdige vesikler, der fremmer inkorporeringen af proteinet i cellemembranen. Det kan også lade sig gøre at anvende vehikler, der inkorporerer proteiner, såsom surfactant protein, såsom SAP(Val) eller SAP(Phe), der udfører denne funktion naturligt, i det mindste for lunge alveolære celler. (PCT Patentansøgning WO/8803170, Whitsett et al, maj 7, 1988 og PCT Patentansøgning W089/04327, Benson et al, mah 18, 1989). De CFTR-holdige vesikler indføres i luftvejene ved inhalation eller irrigation, teknikker, der for tiden anvendes i CF behandling (Boat et al, supra).

### 6.3.2 LÆGEMIDDELTERAPI

Modulation af CFTR funktion kan udføres ved at anvende terapeutiske midler (lægemidler). Disse kan være identificeret ved hjælp af tilfældige approaches under anvendelse af et screeningsprogram, hvor deres effektivitet til modulering af det defekte CFTR protein er monitoreret in vitro. Screeningsprogrammer kan anvende dyrkede celledsystemer, hvor det defekte CFTR protein er udtrykt. Alternativt kan lægemidler designes til at modulere CFTR aktivitet fra viden om struktur- og funktionskorrelationer af CFTR protein og fra viden om den specifikke defekt i de forskellige CFTR mutante proteiner (Capsey and Delvatte, supra). Det er muligt, at hvert mutant CFTR protein vil kræve et anderledes lægemiddel til specifik modulation. Det vil dernæst være nødvendigt for at identificere den(de) specifikke mutation(er) i hver CF patient før initiering af lægemiddelterapi.

Lægemidler kan designes til at interagere med forskellige aspekter af CFTR protein struktur eller funktion. For eksempel kan et lægemiddel (eller et antistof) binde til en strukturel fold af proteinet for at korrigere en defekt struktur. Alternativt vil et lægemiddel kunne binde til en specifik funktionel rest og forøge det affinitet for et substrat eller en cofaktor. Eftersom det er kendt, at medlemmer af klassen af proteiner, hvormed CFTR har strukturel homologi, kan interagere, binde og transportere en række lægemidler, er det rimeligt at forvente, at lægemiddel-relaterede terapier kan være effektive til behandling af CF.

En tredje mekanisme til at forøge aktiviteten af et effektivt lægemiddel ville være at modulere produktionen eller stabiliteten af CFTR inden for cellen. Denne stigning i mængden af CFTR kunne kompensere for dets defekte funktion.

5

Lægemiddelterapi kan også anvendes til at kompensere for den defekte CFTR funktion ved interaktioner med andre komponenter af den fysiologiske eller biokemiske pathway nødvendig for ekspressionen af CFTR funktionen. Disse interaktioner kan føre til stigninger eller formindskelser i aktiviteten af disse afhængige proteiner. Fremgangsmåderne til identifikationen af disse lægemidler ville være svarende til de, der er beskrevet ovenfor for CFTR-relaterede lægemidler.

10

Ved andre genetiske lidelser har det været muligt at korrigere for konsekvenserne af ændret eller manglende normale funktioner ved anvendelse af diæt-modifikationer. Dette har taget form af fjernelse af metabolitter, som når der tale om phenylketonuri, hvor phenylalanin fjernes fra diæten i de første fem leveår med henblik på at forebygge mental retardering, eller ved tilsætningen af store mængder af metabolitter til diæten, som når der er tale om adenosin-deaminase-insufficiens, hvor den funktionelle korrektion af enzymets aktivitet kan frembringes ved tilsætningen af enzymet til diæten. Når således først detaljerne om CFTR funktionen er blevet belyst og den basale defekt i CF er blevet defineret, kan terapi opnås ved diæt-manipulationer.

15

20

Det andet potentielle terapeutiske approach er såkaldt "gen-terapi", hvor normale kopier af CF genet indføres i patienter for således succesfuldt at kode for normalt protein i epitheliale celler med central betydning i berørte væv. Det er mest afgørende at forsøge at opnå dette med luftvejsepitelcellerne i respirationskanalen. CF genet tildeles til disse celler i en form, hvor det kan tages op og kode for tilstrækkeligt protein til at tilvejebringe regulatorisk funktion. Som en følge deraf vil patientens livskvalitet og livslængde være stærkt forbedret. I sidste ende er sigtet naturligvis at tildele genet til alle berørte væv.

30

35

### 6.3.3 GENTERAPI

5 Ét approach til terapi af CF er at indsætte en normal versio... CF genet i luftvejsepitelet fra ramte patienter. Det er vigtigt at bemærke, at respirationssystemet er den primære årsag til morbiditet og mortalitet ved CF; mens pancreatisk sygdom er en væsentlig egenskab, er den relativt velbehandlet i dag med enzym-supplement. Således ville somatisk celle genterapi [for en oversigt, se T. Friedmann, Science 244: 1275 (1989)], der er rettet mod luftvejene lindre de mest alvorlige problemer associeret med CF.

10

A. Retrovirale vektorer. Retrovirus er blevet anset for at være den foretrukne vektor til eksperimenter i somatisk genterapi, med en høj effektivitet af infektion og stabil integration og ekspresion (Orkin et al Prog. Med. Genet 7: 130, (1988)). En mulig ulempe er, at celledeling er nødvendig for retroviral integration, således at de targetede celler i luftvejene kan skulle puffes ind i cellecyklussen forud for retroviral infektion, måske ved hjælp af kemiske midler. Den fulde længde CF gen cDNA kan klones ind i en retroviral vektor og drives fra enten dens endogene promotor eller fra den retrovirale LRT (lang terminal gentagelse (repeat)). Ekspresion af niveauer af det normale protein så lav som 10 % af det endogene mutante protein hos CF patienter ville forventes at være fordelagtigt, eftersom denne er en recessiv sygdom. Tildeling af virusen kunne udføres ved hjælp af aerosol eller instillation i trachea.

15

20

25

B. Andre virale vektorer. Andre tildelingssystemer, der kan være benyttes indbefatter adeno-associeret virus [AAV, McLaughlin et al, J. Virol 62: 1963(1988)], vaccinia virus (Moss et al Annu. Rev. Immunol, 5: 305, 1987)], bovin papillom virus [Rasmussen et al, Methods Enzymol 139: 642 (1987)] eller medlem af herpesvirus gruppen, såsom Epstein-Barr virus (Margolskee et al Mol. Cell. Biol 8: 2937 (1988)). Selv om det vil nødvendigt at lære meget om deres basale biologi, er ideen om at anvende en viral vektor med naturlig tropisme for det respiratoriske træ (e.g. respiratorisk syncytial virus, echovirus, Coxsackie virus etc.) mulig.

30

35

C. Non-viral gen-overførsel. Andre fremgangsmåder til indsætning af CF genet i respiratorisk epitel kan også være produktive; mange af disse har lavere effektivitet og ville potentielt kræve infektion in vitro, selektion af transfektanter

og reimplantation. Dette ville indbefatte calciumphosphat, DEAE dextran, elektroporation og protoplast-fusion. En særligt attraktiv ide er anvendelsen af liposom, der vil kunne være mulig at udføre *in vivo* [Ostro, Liposomes, Marcel Dekker, 1987). Syntetiske kationiske lipider, såsom DOTMA [Felger et al Proc. Natl. Acad. Sci USA 84: 7413 (1987)] kan forøge effektiviteten og fremme udførelse af dette approach.

#### 6.4 CF DYREMODELLER

10 Frembringelsen af en mus eller en anden dyremodel for CF vil være afgørende for at forstå sygdommen og for testning af mulige terapier (for generel oversigt over frembringelse af dyremodeller, se Erickson, Am. J. Hum. Genet 43: 582 (1988)). For tiden eksisterer der ingen dyremodel for CF. Den

15 evolutionsmæssige konservering af CF genotyper (som påvist ved hjælp af tværart hybridisering blots for E4.3 og H1.6), hvilket er vist i figur 4, viser, at et ortologt gen eksisterer i musen (skal herefter betegnes mCF, og dets korresponderende protein som mCFTR), og dette vil være muligt at klonere i muse genomiske og cDNA biblioteker under anvendelse af de humane CF gen-prober. Det forventes, at frembringelsen af en specifik mutation i muse

20 genet analogt med F508 mutationen vil være mest optimal til at reproducere fænotypen, selv om komplet inaktivering af mCFTR genotyper også vil være en anvendelig mutant at frembringe.

A. Mutagenese. Inaktivering af mCF genotyper kan opnås ved hjælp af kemikalier (e.g. Johnson et al Proc. Natl. Acad. Sci. USA 78:3138 (1981)) eller røntgen mutagenese [Popp et al J. Mol. Biol. 127: 141 (1979)] af muse gameter, efterfulgt af fertilisering. Afkom-heterozygoter for inaktivering af mCFTR kan dernæst identificeres ved hjælp af Southern blotting for at påvise tab af én allel ved dosering, eller svigt i at nedarve én parental allel, hvis en RFLP markør fastsættes. Dette approach er tidligere blevet succesfuldt anvendt til at

30 identificere muse mutanter for  $\alpha$ -globin [Whitney et al Proc. Natl. Acad. Sci. USA 77: 1087 (1980)], phenylalaninhydroxylase [McDonald et al, Pediatr. Res 23: 63 (1988)], og carbonanhydrase II (Lewis et al Proc. Natl. Acad. Sci. USA 85: 1962, (1988)).

35

B. Transgen teknologi En normal eller mutant version af CFTR eller mCFTR kan indsættes i muse kimplacentalinjen under anvendelse af nu standard teknikker



for oocyt-injektion [Camper, Trends in Genetics (1988)]; hvis alternativt det er ønskeligt at inaktivere eller erstatte det endogene mCF gen, kan det homologe rekombinationssystem under anvendelse af embryonale stam (ES) celler [Capecchi, Science 244: 1288 (1989)] anvendes.

5

1. Oocyt-injektion Anbringelse af én eller flere kopier af det normale eller mutante mCF gen ved en tilfældig lokalisering i muse kimcellelinjen kan udføres ved hjælp af mikroinjektion af pro-kernen fra en nyligt fertiliseret muse oocyt, efterfulgt af reimplantation i en pseudo-gravid foster moder. De levendefødte mus kan dernæst screenes for integranter under anvendelse af analyse af hale DNA for tilstedeværelsen af humane CF gen sekvenser. Den samme protokol kan anvendes til at indsætte et mutant mCF gen. For at frembringe en muse model, ville man ønske at anbringe dette transgen i en muse baggrund, hvor det endogene mCF gen er blevet inaktiveret, enten ved hjælp af mutagenese (se ovenfor) eller ved hjælp af homolog rekombination (se nedenfor). Den transgene kan være enten: a) en komplet genomisk sekvens, selv om størrelsen af denne (omtrentligt 250 kb) ville kræve, at den injiceres som et gær kunstigt kromosom eller et kromosomfragment; b) en cDNA med enten den naturlige promotor eller en heterolog promotor; c) et "minigen", der indeholder hele den kodende region og forskellige andre elementer, såsom introns, promotor og 3' flankerende elementer, der har vist sig at være nødvendige for optimal ekspression.

10

15

20

## 2. Retroviral infektion af tidligere fostre.

25

Dette alternativ involverer indsætning af CFTR eller mCF genet i en retroviral vektor og direkte inficering af muse-fostre på tidligere stadier af udvikling til frembringelse af en kimære [Soriano et al Cell 46: 19 (1986)]. I det mindste vil nogle af disse føre til kimcellelinje-transmission.

30

## 3. ES celler og homolog rekombination.

35

Det embryonale stamcelle approach [Capecchi, supra og Capecchi, Trends Genet 5:70(1989)] tillader muligheden for at udføre gen-overførsel og dernæst screening af de dannede totipotente celler med henblik på at identificere de sjældne homologe rekombinationsbegivenheder. Når de først er identificeret, kan disse anvendes til at frembringe kimærer ved injektion af muse

blastocyster, og en proportion af de dannede mus vil udvise kimcellelinje transmission fra den rekombinante linje. Der er adskillige måder, hvorpå denne kunne være anvendelige i frembringelsen af en muse model for CF:

- 5 a) Inaktivering af mCF genet kan hensigtsmæssigt udføres ved at designe et DNA fragment, der indeholder sekvenser fra en mCFTR exon, som flankerer en selekterbar markør, såsom neo. Homolog rekombination vil føre til insertion af neo sekvenserne i midten af en exon, inaktivering af mCFTR. De homologe rekombinationsbegivenheder (sædvanligvis omtrentligt 1 ud af 1000) kan
- 10 genkendes fra de heterologe ved hjælp af DNA analyse af individuelle kloner (sædvanligvis under anvendelse af PCR, Kim et al Nucleic Acids Res. 16:8887 (1988), Joyner et al Nature 338:153 (1989); Zimmer et al supra, p. 150) eller ved at anvende en negativ selektion over for de heterologe begivenheder (såsom anvendelsen af et HSV TK gen ved enden af konstruktionen, efterfulgt
- 15 af gancyclovir selektion, Mansour et al, Nature 336:348 (1988)]. Denne inaktiverede mCFTR mus kan dernæst anvendes til at introducere et mutant CF gen eller mCF gen, der indeholder F508 abnormaliteten eller en hvilken som helst anden ønsket mutation.
- 20 b) Det er muligt, at specifikke mutanter af mCFTR kan frembringe i ét trin. For eksempel kan man frembringe en konstruktion, der indeholder mCF intron 9 sekvenser ved 5' enden, et selekterbart neo gen i midten, og intro 9 + exon 10 (der indeholder muse versionen af F508 mutationen) ved 3' enden. En homolog rekombinationsbegivenhed ville føre til insertionen af neo genet i
- 25 intron 9 og erstatningen af exon 10 med den mutante version.
- c) Hvis tilstedeværelsen af den selekterbare neo markør i intronen ændrede ekspresion af mCF genet, ville det være muligt at udskære den i et andet homologt rekombinationstrin.
- 30 d) Det er også muligt at frembringe mutationer i muse kimcellelinjen ved at injicere oligonucleotider, der indeholder den mutation, som har interesse, og screening af de dannede celler ved hjælp af PCR.
- 35 Denne udførelsesform ifølge den foreliggende opfindelse har primært betragtet en muse-model for cystisk fibrose. Figur 4 viser tvær-art hybridisering ikke kun til muse DNA, men også til bovin, hamster og kylling DNA. Det overvejes

således, at et ortologt gen også vil eksistere i mange andre arter. Det overvejes således, at det vil være muligt at frembringe andre dyremodeller under anvendelse af lignende teknologi.

Patentkrav

1. Oprensat DNA molekyle, der omfatter et cystisk fibrose gen, idet nævnte gen omfatter en DNA sekvens udvalgt fra gruppen bestående af:

5

(a) DNA sekvenser, der svarer til DNA sekvensen i figur 1 fra aminosyrerestposition 1 til position 1480 eller en polymorf form af nævnte DNA sekvens, hvor nucleotidsubstitution i nævnte DNA sekvens ikke påvirker den essentielle funktion af polypeptidet kodet deraf;

10

(b) DNA sekvenser, der koder et normalt cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid med sekvensen ifølge figur 1 fra aminosyrerestposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydrofatiske profil eller sekundære eller tertiære struktur;

15

(c) DNA sekvenser, som svarer til et fragment af DNA sekvensen i figur 1, herunder mindst 16 sekventielle nucleotider mellem aminosyrerestposition 134 og 4573, og som kan screene for og påvise tilstedeværelsen af enten et normalt eller defekt cystisk fibrose gen eller genprodukt;

20

(d) DNA sekvenser, der omfatter mindst 16 nucleotider og koder et fragment af aminosyresekvensen i figur 1, som kan screene for og påvise tilstedeværelsen af enten et normalt eller defekt cystisk fibrose gen eller genprodukt; og

25

(e) DNA sekvenser, der koder en epitop kodet af mindst 18 sekventielle nucleotider i DNA sekvensen i figur 1 mellem aminosyrerestposition 1 og 1480.

30

2. Oprensat DNA molekyle (a), (b) eller (c) ifølge krav 1, hvilket DNA molekyle yderligere indeholder mindst én mutation i den DNA sekvens, der, hvis den udtrykkes i cellerne i det menneskelige legeme, er associeret med ændret cellefunktion, som korrelerer med cystisk fibrose, hvor nævnte mutation indbefatter en deletion af tre nucleotider, der koder phenylalanin ved aminosyrerestposition 508.

35

3. Oprensat nucleinsyreprobe, der omfatter en DNA eller RNA nucleotid sekvens svarende til sekvensen ifølge egenskab (c), (d) eller (e) ifølge krav 1 eller 2.
- 5 4. Nucleinsyreprobe ifølge krav 3, hvor nævnte sekvens omfatter AAA GAA AAT ATC ATC TTT GGT GTT og dens komplement.
5. Rekombinant kloningsvektor, der omfatter DNA molekylet ifølge krav 1 eller 2.
- 10 6. Non-human vært transformeret med vektoren ifølge krav 5.
7. Fremgangsmåde til fremstilling af et normalt eller mutant cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid, hvilken fremgangsmåde omfatter trinene med:
- 15 (a) dyrkning af en værtselle transficeret med vektoren ifølge krav 5 i et medium og under betingelser, der er favorable for ekspresion af normalt eller mutant CFTR polypeptid; og
- 20 (b) isolering af det udtrykte normale eller mutante CFTR polypeptid.
8. Oprensat DNA molekyle, der omfatter en DNA sekvens, som koder et mutant cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid med sekvensen ifølge figur 1 for aminosyrerestposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur, hvilket DNA molekyle yderligere er kendetegnet ved en deletion af tre nucleotider, der koder phenylalanin ved aminosyrerestposition 508.
- 25 30 35 9. Oprensat cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid kendetegnet ved at have en peptid-molekylvægt på 170.000 dalton og udvalgt fra gruppen bestående af:

- 5 (a) et normalt CFTR polypeptid med celle transmembran ion konduktans påvirkende aktivitet og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte normale CFTR polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydropatiske profil eller sekundære eller tertiære struktur; og
- 10 (b) et mutant CFTR polypeptid med cystisk fibrose aktivitet i humane celler og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyrerestposition 508.
- 15 10. Polypeptid kodet af ekspresion af en DNA sekvens ifølge krav 1 eller 2, idet nævnte polypeptid udviser den immunologiske aktivitet af et normalt eller mutant cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid.
- 20 11. Fremgangsmåde til at isolere det normale eller mutante CFTR protein ifølge krav 10 fra celler, der indeholder nævnte protein, hvilken fremgangsmåde omfatter trinene med:
- 25 (a) opløsning af protein fra en selekteret cellemembran, hvor enten nævnte normale eller mutante CFTR protein er udtrykt, med henblik på at tilvejebringe en opløsning af nævnte CFTR protein:
- 30 (b) adskillelse af nævnte CFTR protein fra nævnte opløsning ved at kontakte nævnte opløsning med antistoffer mod nævnte normale eller mutante CFTR protein, idet nævnte antistoffer er immobiliseret på et substrat;
- 35 (c) skylning af nævnte substrat med henblik på at fjerne protein, der ikke klæbede til nævnte antistoffer;
- (d) frigivelse af nævnte CFTR protein fra nævnte antistoffer for derved at isolere nævnte CFTR protein, og
- (e) oprensning af nævnte CFTR protein for at fjerne eventuelt resterende andet mammalt protein.

12. Fremgangsmåde til at screene et individ med henblik på at bestemme, om nævnte individ er en cystisk fibrose (CF) bærer eller en CF patient, hvilken fremgangsmåde omfatter:

5 at tilvejebringe et assay til påvisning i en biologisk prøve, der skal screenes, af tilstedeværelsen af mindst et medlem fra gruppen bestående af et normalt CF gen, normale CF genprodukter, et mutant CF gen, mutant CF genprodukter og blandinger deraf, et normalt CFTR polypeptid med celle transmembran ion  
10 konduktans påvirkende aktivitet og med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte normale CFTR polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf, eller dets hydropatiske profil eller sekundære eller tertiære struktur, et mutant CFTR polypeptid med cystisk fibrose aktivitet i humane celler og med  
15 sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyrestposition 508, hvor assayet er

(i) et assay, der involverer hybridisering, eller

20 (ii) et immunologisk assay,

og/eller er opnået ved at anvende én eller flere hybridiseringer under  
anvendelse af specifikke oligonucleotider, direkte DNA sekventering,  
25 restriktionsenzymmer, skelnen på basis af elektroforetisk mobilitet i geler med denaturerende middel, RNase beskyttelse, kemisk spaltning eller den ligase-medierede påvisningsprocedure.

13. Fremgangsmåde ifølge krav 12, hvor assayet er et assay, der involverer hybridisering, og den biologiske prøve indbefatter mindst en del af individets  
30 genom.

14. Fremgangsmåde ifølge krav 13, hvor assayet yderligere omfatter en mærket nucleotidprobe ifølge krav 3.

35 15. Fremgangsmåde ifølge krav 14, hvor nævnte probe omfatter nucleotid-sekvensen ifølge krav 4.

16. Fremgangsmåde ifølge krav 12, hvor assayet er et immunologisk assay, og den biologiske prøve indbefatter et normalt cystisk fibrose transmembrane konduktans regulator (CFTR) polypeptid fra individet og/eller et mutant CFTR polypeptid fra individet: nævnte normale CFTR polypeptid med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydrofatiske profil eller sekundære eller tertiære struktur; og nævnte mutante CFTR polypeptid med sekvensen ifølge figur 1 for aminosyreposition 1 til 1480, bortset fra deletionen af phenylalanin ved aminosyreposition 508.
17. Fremgangsmåde ifølge krav 16, hvor assayet yderligere indbefatter et antistof specifikt for et normalt CFTR polypeptid.
18. Fremgangsmåde ifølge krav 16, hvor assayet yderligere indbefatter et antistof specifikt for et mutant CFTR polypeptid.
19. Fremgangsmåde ifølge krav 17 eller 18, hvor antistoffet er mindst ét monoklonalt antistof.
20. Fremgangsmåde ifølge krav 12, hvor individet er et humant foster in utero.
21. Fremgangsmåde ifølge krav 14 eller 15, hvor assayet yderligere indbefatter mindst én yderligere nucleotidprobe ifølge krav 3, under forudsætning af at hver nucleotidprobe i nævnte assay er forskellig.
22. Fremgangsmåde til screening for en potentiel cystisk fibrose (CF) bærer eller patient med henblik på at identificere tilstedeværelsen af en identificeret cystisk fibrose mutation i det CF gen, der koder et cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid med sekvensen ifølge figur 1 fra aminosyreposition 1 til position 1480 eller en polymorf form af nævnte polypeptid, hvor aminosyresubstitution i de variable regioner af nævnte polypeptid med sekvensen ifølge figur 1 ikke påvirker den essentielle funktion deraf eller dets hydrofatiske profil eller sekundære eller tertiære struktur, idet fremgangsmåden indbefatter trinene med:



(a) hybridisering af en DNA probe ifølge krav 3 på isoleret genomisk DNA fra nævnte potentielle CF bærer eller nævnte potentielle patient, idet nævnte DNA probe udspænder nævnte cystisk fibrose mutation i nævnte CF gen, hvor nævnte DNA probe er i stand til at påvise nævnte cystisk fibrose mutation; og

5

(b) behandling af nævnte genomiske DNA med henblik på at bestemme tilstedeværelsen af nævnte DNA probe og derved vise, ifølge en forud bestemt måde for hybridisering, tilstedeværelsen eller fraværet af nævnte cystisk fibrose mutation.

10

23. Fremgangsmåde til påvisning af cystisk fibrose bærere eller patienter, hvor nævnte fremgangsmåde består i at kontakte et CF gen fra patienten eller bæreren med en restriktionsendonuclease, at bestemme tilstedeværelsen eller fraværet af et restriktionsendonuclease-site i CF genet og at sammenligne mønsteret for restriktionsendonuclease-sites i CF genet med et normalt CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerest ved position 1 og 1480 i figur 1, idet tilstedeværelsen af sites foruden det normale CF gen eller fraværet af sites i det normale CF gen er indikativt for et mutant CF gen.

15

20

24. Fremgangsmåde til påvisning af cystisk fibrose bærere, hvor nævnte fremgangsmåde består i at bestemme differential mobilitet af heteroduplex PCR produkter i polyacrylamidgeler som et resultat af insertioner eller deletioner i det mutante CF gen sammenlignet med det normale CF gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerest ved position 1 og 1480 i figur 1.

25

25. Kit til at undersøge for tilstedeværelsen af et normalt cystisk fibrose (CF) gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerestposition 1 og 1480 i figur 1 eller et mutant CF gen, der har mindst en tre basepar deletion i det normale CF gen, hvilken deletion resulterer i deletionen af en phenylalanin fra aminosyrerestposition 508 i figur 1, ved hjælp af immunoassay, hvilket kit omfatter:

30

35

(a) et antistof, der specifikt binder til et genprodukt af det normale eller det mutante CF gen ifølge krav 9;

(b) reagensmidler til påvisning af antistoffets binding til genproduktet; og

(c) idet antistoffet og reagensmidlerne hver er til stede i mængder, der er effektive til at udføre immunoassayet.

5

26. Kit til undersøgelse for tilstedeværelsen af et normalt cystisk fibrose (CF) gen svarende til den nucleotid-sekvens, der koder polypeptidet mellem aminosyrerestposition 1 og 1480 i figur 1 eller et mutant CF gen med mindst en tre basepar deletion i det normale CF gen, hvilken deletion resulterer i deletionen af en phenylalanin fra aminosyrerestposition 508 i figur 1, ved hjælp af hybridisering, hvilket kit omfatter:

10

(a) en oligonucleotidprobe ifølge krav 3, hvilken oligonucleotidprobe specifikt binder til det normale eller det mutante CF gen;

15

(b) reagensmidler til påvisning af hybridiseringen af oligonucleotidproben til det normale eller det mutante CF gen; og

(c) idet proben og reagensmidlerne hver er til stede i mængder effektive til at udføre hybridiseringsassayet.

20

27. Immunologisk aktivt anti-cystisk fibrose transmembran konduktans regulator (CFTR) polyklonalt eller monoklonalt antistof specifikt for et CFTR polypeptid ifølge krav 9.

25

28. Hybridom, der producerer et monoklonalt antistof specifikt for et cystisk fibrose transmembran konduktans regulator (CFTR) polypeptid ifølge krav 9.

29. Sammensætning til behandlingen af cystisk fibrose i en patient, hvilken sammensætning omfatter en terapeutisk effektiv mængde af proteinet ifølge egenskab (a) ifølge krav 9 i kombination med en farmaceutisk acceptabel bærer.

30

30. Sammensætning ifølge krav 29, hvor nævnte bærer indbefatter et lunge overfladeaktivt protein for at fremme påføring af nævnte sammensætning til respiratoriske epitelceller.

35

31. Sammensætning til genterapi af cystisk fibrose, hvilken sammensætning omfatter et DNA molekyle ifølge krav 1 og en vehikel til tildeling af nævnte DNA molekyle til en celle af en cystisk fibrose patient.
- 5 32. Sammensætning ifølge krav 31, hvor nævnte vehikel er en rekombinant vektor.
33. Non-humant dyr, der omfatter et heterologt cellesystem, som omfatter en rekombinant kloningsvektor ifølge krav 5, hvilken vektor inducerer cystisk fibrose symptomer i nævnte dyr.
- 10
34. Transgen mus, der udviser cystisk fibrose symptomer, hvilken transgen mus er opnåelig ved hjælp af de trin, der omfatter indføring af et cystisk fibrose gen ifølge krav 2, hvilket gen koder for et mutant cystisk fibrose transmembran konduktans regulator protein, der indbefatter mindst en deletion af phenylalanin ved den position, der svarer til 508 i figur 1 til en muse-oocyt eller et muse-foster; og eventuelt at inaktivere det endogene muse cystisk fibrose gen.
- 15

FIG.1.

1 AATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCA  
 61 GAGTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGC  
 121 GCCCAGAGACCATGCAGAGGTCGCTCTGGAAAAGCCAGCGTTGTCTCCAACTTTT  
 M O R S P L E K A S V V S K L F 16  
 181 TTCACCTGGACCAGACCAATTTTGAGGAAAGGATACAGACAGCGCCTGGAATTGTCAGAC  
 F S W T R P I L R K G Y R Q R L E L S D 36  
 241 ATATACCAAATCCCTTCTGTTGATTCTGCTGACAACTATCTGAAAAATTGGAAADAGAA  
 I Y O I P S V D S A D N L S E K L E H E 56  
 301 TGGGATAGAGAGCTGGCTTCAAAGAAAAATCCTAAACTCATTAAATGCCCTTCGGCGATGT  
 W D R E L A S K K N P K L I N A L R R C 76  
 361 TTTTCTGGAGATTATGTTCTATGGAATCTTTTATATTTAGGGGAAGTCACCAAAGCA  
 F F W R F M P Y G I F L Y L G E V T K A 96  
 421 GTACAGCCTCTCTTACTGGGAAGAATCATAGCTTCTATGACCCGGATAACAAGGAGGAA  
 V O P L L L G R I I A S Y D P D N K E E 116  
 481 CGCTCTATCGGATTTATCTAGGCATAGGCTTATGCCCTTCTTTATTGTGAGGACACTG  
 R S I A I Y L G I G L C L L F I V R T L 136  
 541 CTCCTACACCCAGCCATTTTGGCCTTCATCACATTGGAATGCAGATGAGAATAGCTATG  
 L L H P A I F G L H H I G M O M R I A H 156  
 601 TTTAGTTTGATTATAAGAAGACTTTAAAGCTGCAAGCCGTGTTCTAGATAAAAATAAGT  
 F S L I Y K K T L K L S S R V L D K I S 176  
 661 ATTGGACAACCTTGTTAGTCTCCTTCCAACAACCTGAACAAATTGATGAAGGACTTGCA  
 I G O L V S L L S N N L N K F D E G L A 196  
 721 TTGGCACATTTCTGTGGATCGCTCCTTGGCAAGTGGCACTCCTCATGGGGCTAATCTGG  
 L A H F V W I A P L O V A L L M G L I W 216  
 781 GAGTTGTTACAGGCGTCTGCCTTCTGTGGACTTGGTTTCTGATAGTCTTGCCTTTTT  
 E L L O A S A F C G L G F L I V L A L F 236  
 841 CAGGCTGGGCTAGGGAGAATGATGATGAAGTACAGAGATCAGAGAGCTGGGAAGATCAGT  
 O A G L G R M H M K Y R D O R A G K I S 256  
 901 GAAAGACTTGTGATTACCTCAGAAATGATTGAAAATATCCAATCTGTTAAGGCATACTGC  
 E R L V I T S E M I E N I O S V K A Y C 276  
 961 TGGGAAGAAGCAATGGAAAAATGATTGAAAACCTAAGACAAACAGAACTGAAACTGACT  
 W E E A M E K M I E N L R Q T E L K L T 296  
 1021 CCGAAGGCAGCCTATGTGAGATACTTCAATAGCTCAGCCTTCTTCTCAGGGTTCTTT  
 R K A A Y V R Y F N S S A F F F S G F F 316  
 1081 GTGGTGTTTTTATCTGTGCTTCCCTATGCACTAATCAAAGGAATCATCCTCCGGAAATA  
 V V F L S V L P Y A L I M G I I L R K I 336  
 1141 TTCACCACCATCTCATTCTGCATTGTTCTGCGCATGGCGGTCACTCGGCAATTCCTCTGG  
 F T T I S F C I V L R M A V T R Q F P W 356  
 1201 GCTGTACAAACATGGTATGACTCTCTGGAGCAATAACAAAATACAGGATTTCTTACAA  
 A V Q T W Y D S L G A I N K I Q D F L O 376  
 1261 AAGCAAGAAATAAGACATTGGAATATACTTACGACTACAGAAGTAGTGATGGAGAAT  
 K Q E Y K T L E Y N L T T T E V V M E N 396  
 1321 GTAACAGCCTTCTGGGAGGAGGATTGGGGAATTATTTGAGAAAGCAAACAAAACAAT  
 V T A F W E E G F G E L F E K A K O N N 416

FIG.1(cont'd)

1381 N N R K T S N G D D S L F F S N F S L L L 436  
 AACAAATAGAAAACTTCTAATGGTGATGACAGCCTCTTCTTCAGTAATTTCTCACTTCTT  
 1441 G T P V L K D I N F K I E R G Q L L A V 456  
 GGTACTCCTGTCTGAAAGATATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTT  
 1501 A G S T G A G K T S L L M N I M G E L E 476  
 GCTGGATCCACTGGAGCAGGCAAGACTTCACTTCTAATGATGATTATGGGAGAACTGGAG  
 1561 P S E G K I K H S G R I S F C S O F S W 496  
 CCTTCAGAGGGTAAAATTAAGCACAGTGGAAAGAAATTTCACTTCTGTTCTCAGTTTTCTGG  
 1621 I M P G T I K E N I I F G V S Y D E Y R 516  
 ATTATGCCTGGCACCATTAAAGAAAAATATCATCTTTGGTGTTCCTATGATGAATATAGA  
 1681 Y R S V I K A C Q L E E D I S K F A E K 536  
 TACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGACATCTCCAAGTTTGCAGAGAAA  
 1741 D N I V L G E G G I T L S G G Q R R A R I 556  
 GACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATT  
 1801 S L A R A V Y K D A D L Y L L D S P F G 576  
 TCTTTAGCAAGGCGAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGG  
 1861 Y L D V L T E K E I F E S C V C K L M A 596  
 TACCTAGATGTTTAAACAGAAAAAGAAATATTTGAAAGCTGTGTCTGTAAACTGATGGCT  
 1921 N K T R I L V T S K M E H L K K A D K I 616  
 AACAAAACCTAGGATTTTGGTCACTTCTAAAATGGAACATTTAAAGAAAGCTGACAAAATA  
 1981 L I L N E G S S Y F Y G T F S E L O N L 636  
 TTAATTTTGAATGAAGGTAGCAGCTATTTTATGGGACATTTTTCAGAACTCCAAAATCTA  
 2041 O P D F S S K L H G C D S F D Q F S A E 656  
 CAGCCAGACTTTAGCTCAAAACTCATGGGATGTGATTCTTTTCGACCAATTTAGTCAGAA  
 2101 R R N S I L T E T L H R F S L E G D A P 676  
 AGAAGAAATTCATCCTAAGTACACTGAGACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCT  
 2161 V S W T E T K K O S F K Q T G E F G E K 696  
 GTCCTCTGGACAGAAACAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGGGAAAAA  
 2221 R K N S I L N P I N S I R K F S I V O K 716  
 AGGAAGAATTCTATTCTCAATCCAATCAACTCTATACGAAAATTTCCATTGTGCAAAAG  
 2281 T P L Q M N G I E E D S D E P L E R R L 736  
 ACTCCCTTACAAATGAATGGCATCGAAGAGGATTCTGATGAGCCTTTAGAGAGAAGGCTG  
 2341 S L V P D S E O G E A I L P R I S V I S 756  
 TCCTTAGTACCAGATTCTGAGCAGGGAGAGGGGATACTGCCTCGCATCAGCGTGATCAGC  
 2401 T G P T L Q A R R R O S V L N L M T H S 776  
 ACTGGCCCCACGCTTCAGGCACGAAGGAGGCGTCTGTCCTGAACCTGATGACACACTCA  
 2461 V N Q G Q N I H R K T T A S T R K V S L 796  
 GTTAACCAAGGTCAGAACATTACCCGAAAGACAACAGCATCCACAGGAAAAGTGTCACTG  
 2521 A P O A N L T E L D I Y S R R L S O E T 816  
 GCCCCTCAGGCAAACTTGACTGAAGTATATATATTCAAGAAGGTTATCTCAAGAAAAT

FIG.1 (cont'd)

2581 G L E I S E E I N E E D L N E C L F D D 836  
 GGCTTGGAAATAAGTGAAGAAATTAACGAAGAAGACTTAAAGGAGTGCCTTTTTGATGAT

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

2701 K S L I F V L I W C L V I F L A E V A A 876  
 AAGAGCTTAATTTTTGTGCTAATTTGGTGTAGTAATTTTCTGGCAGAGGTGGCTGCT

2761 S L V V L W L L G M T P L Q D K G N S T 896  
 TCTTTGGTTGTGCTGTGGCTCCTTGGAAACTCCTCTTCAAGACAAAGGGATAGTACT

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

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

2941 L V H T L I T V S K I L H H K M L H S V 956  
 CTGGTGCATACTCTAATCACAGTGTGAAAATTTACACCACAAAATGTTACATTCTGTT

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

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

3121 L L L I V I G A I A V V A V L Q P Y I F 1016  
 TTGTTATTAATTTGTGATTGGAGCTATAGCAGTTGTGCGCAGTTTTACAACCCTACATCTTT

3181 V A T V P V I V A F I M L R A Y F L Q T 1036  
 GTTGCACAGTCCAGTGATAGTGGCTTTTATTATGTTGAGAGCATATTTCTCCAAACC

3241 S Q O L K O L E S E G R S P I F T H L V 1056  
 TCACAGCAACTCAAACAAGTGAATCTGAAGGCAGGAGTCCAATTTTCACTCATCTTGT

3301 T S L K G L N T L R A F G R O P Y F E T 1076  
 ACAAGCTTAAAAGGACTATGGACACTTCGTGCCTTCGGACGGCAGCCTTACTTTGAAACT

3361 L F H K A L N L H T A N W F L Y L S T L 1096  
 CTGTTCCACAAAGCTCTGAATTTACATACTGCCAACTGGTCTTGTACCTGTCAACACTG

3421 R N F Q M R I E M I F V I F F I A V T F 1116  
 CGCTGGTTCCAAATGAGAATAGAAATGATTTTGTGATCTTCTTCATTGCTGTTACCTTC

3481 I S I L T T G E G E G R V G I I L T L A 1136  
 ATTTCCATTTTAAACAAGGAGAAGGAAGGAGTGGTATTATCCTGACTTTAGCC

3541 M N I M S T L O W A V N S I D V D S L 1156  
 ATGAATATCATGAGTACATTGCAGTGGCTGTAAACTCCAGCATAGATGTGGATAGCTTC

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

3661 K S T K P Y K N G Q L S K V M I I E N S 1196  
 AAGTCAACCAAACCATACAAGAATGCCAACTCTCGAAAGTTATGATTATTGAGAATTCA

3721 H V K K D D I W P S G G O M T V K D L T 1216  
 CACGTGAAGAAAGATGACATCTGGCCCTCAGGGGGCCAAATGACTGTCAAAGATCTCACA

3781 A K Y T E G G N A I L E N I S F S I S P 1236  
 GCAAATACACAGAAGGTGGAAATGCCATATTAGAGAACATTTCTTCTCAATAAGTCTT

3841 G O R V G L L G R T G S G K S T L L S A 1256  
 GGCCAGAGGTTGGCCTCTTGGGAAGAAGTGGATCAGGGAAGAGTACTTTGTTATCAGCT

FIG.1. (cont'd)

3901 F L R L L N T E G E I Q I D G V S W D S 1276  
 TTTTGAGACTACTGAACACTGAAGGAGAAATCCAGATCGATGGTGTGTCTTGGGATTCA

3961 I T L Q Q W R K A F G V I P Q K V F I F 1296  
 ATAACCTTGCAACAGTGGAGGAAAGCCTTTGGAGTGATACCACAGAAAGTATTTATTTTT

4021 S G T F R K N L D P Y Z O W S D O E I W 1316  
 TCTGGAACATTTAGAAAAA<sup>o</sup>ACTTGGATCCCTATGAACAGTGGAGTGATCAAGAAATATGG

4081 K V A D E V G L R S V I E O F P G K L D 1336  
 AAAGTTGCAGATGAGTTGGGCTCAGATCTGTGATAGAACAGTTTCCTGGGAAGCTTGAC

4141 F V L V D G G C V L S E G H K O L M C L 1356  
 TTTGTCTTGTGGATGGGGGCTGTGTCTTAAGCCATGGCCACAAGCAGTTGATGTGCTTG

4201 A R S V L S K A K I L L L D E P S A H L 1376  
 GCTAGATCTGTTCTCAGTAAGGCGAAGATCTTGGCTGCTGATGAACCCAGTGCTCATTG

4261 D P V T Y Q I I R R T L K O A F A D C T 1396  
 GATCCAGTACATAACCAATAATTAGAAGAACTCTAAAACAAGCATTGCTGATTGCACA

4321 V I L C E H R I E A M L E C O O F L V I 1416  
 GTAATTCTCTGTGAACACAGGATAGAAGCAATGCTGGAATGCCAACAATTTTTGTGCATA

4381 E E N K V R O Y D S I O K L L N E R S L 1436  
 GAAGAGAACAAGTGGGGCAGTACGATTCCATCCAGAAACTGCTGAACGAGAGGAGCCTC

4441 F R O A I S P S D R V K L F P H R N S S 1456  
 TTCCGGCAAGCCATCAGCCCCTCCGACAGGGTGAAGCTCTTCCCCACCGGA<sup>o</sup>ACTCAAGC

4501 K C K S K P Q I A A L K E E T E E E V Q 1476  
 AAGTGAAGTCTAAGCCCCAGATTGCTGTCTGAAAGAGGAGACAGAAGAAGAGGTGCAA

D T R L = 1480

4561 GATACAAGGCTTTAGAGAGCAGCATAAATGTTGACATGGGACATTTGCTCATGGAATTGG

4621 AGCTCGTGGGACAGTCACCTCATGGAATTGGAGCTCGTGGAACAGTTACCTCTGCCTCAG

4681 AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAAGAAACATTTGGTAAGGGGAATTGAGG

4741 ACACTGATATGGGTCTTGATAAATGGCTTCTGGCAATAGTCAAATGTGTGAAAGGTAC

4801 TTCAAATCCTTGAAGATTTACCCTTGTGTTTTGCAAGCCAGATTTTCTGAAAACCTT

4861 GCCATGTGCTAGTAATTGGAAAGGCAGCTCAAATGTCAATCAGCCTAGTTGATCAGCTT

4921 ATTGTCTAGTGAAACTCGTTAATTTGTAGTGTGGAGAACAAC<sup>o</sup>TCAAATCATACTTCTTA

4981 GGGTTATGATTAAGTAATGATAACTGGAAACTTCAGCGGTTTATATAAGCTTGTATTCTT

5041 TTTCTCTCCTCTCCCCATGATGTTTAGAAACACAATATATTGTTTGCTAAGCATTCCA

5101 ACTATCTCATTTCCAAGCAAGTATTAGAATACCACAGGAACCACAAGACTGCACATCAAA

5161 ATATGCCCCATTCAACATCTAGTGAGCAGTCAGGAAAGAGAACTTCCAGATCCTGGAAAT

5221 CAGGGTAGTATTGTCCAGGTCACCAAAAATCTCAATATTTTCAGATAATCAATACAT

5281 CCCTTACCTGGGAAAGGGCTGTTATAATCTTTCACAGGGGACAGGATGGTTCCCTTGATG

5341 AAGAAGTTGATATGCCTTTTCCCAACTCCAGAAAGTGACAAGCTCACAGACCTTTGAACT

5401 AGAGTTTAGCTGGAAAAGTATGTTAGTGCAAAATGTCACAGGACAGCCCTTCTTCCACA

5461 GAAGCTCCAGGTAGAGGGTGTGTAAGTAGATAGGCCATGGGCACTGTGGGTAGACACACA

5521 TGAAGTCCAAGCATTAGATGTATAGGTTGATGGTGGTATGTTTTTCAGGCTAGATGTATG

5581 TACTTCATGCTGTCTACACTAAGAGAGAATGAGAGACACACTGAAGAAGCACC<sup>o</sup>AATCATG

5641 AATTAGTTTTATATGCTTCTGTTTTATAATTTTGTGAAGCAAAATTTTTCTCTAGGAAA

5701 TATTTATTTAATAATGTTCAAACATATATTACAATGCTGTATTTTAAAAGAATGATTA

5761 TGAATTACATTTGTATAAAAATAATTTTATATTTGAAATATTGACTTTTTATGGCACTAG

5821 TATTTTATGAAATATTATGTTAAAAC<sup>o</sup>TGGGACAGGGGAGAACC<sup>o</sup>TAGGGTGTATTAACC

5881 AGGGCCATGAATCACCTTTGGCTCGGAGGGAAGCCTTGGGCTGATCGAGTTGTTGCC

5941 CACAGCTGTATGATTCCAGCCAGACACAGCCTCTTAGATGCAGTTCTGAAGAAGATGGT

6001 ACCACCAGTCTGACTGTTTCCATCAAGGGTACACTGCCTTCTCAACTCAA<sup>o</sup>ACTGACTCT

6061 TAAGAAGACTGCATTATATTTACTGTAAGAAAAATATCACTTGTCAATAAAATCCATA

6121 CATTGTGT (A) n





FIG.2. (cont'd)

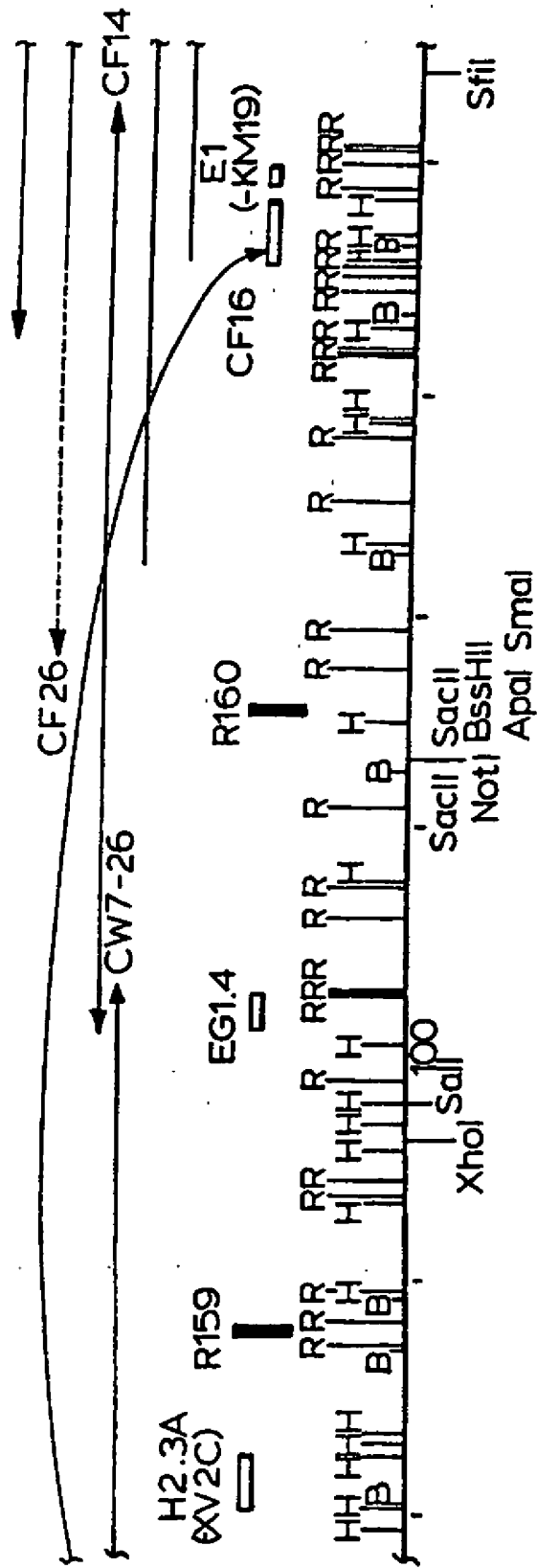


FIG.2 (cont'd)

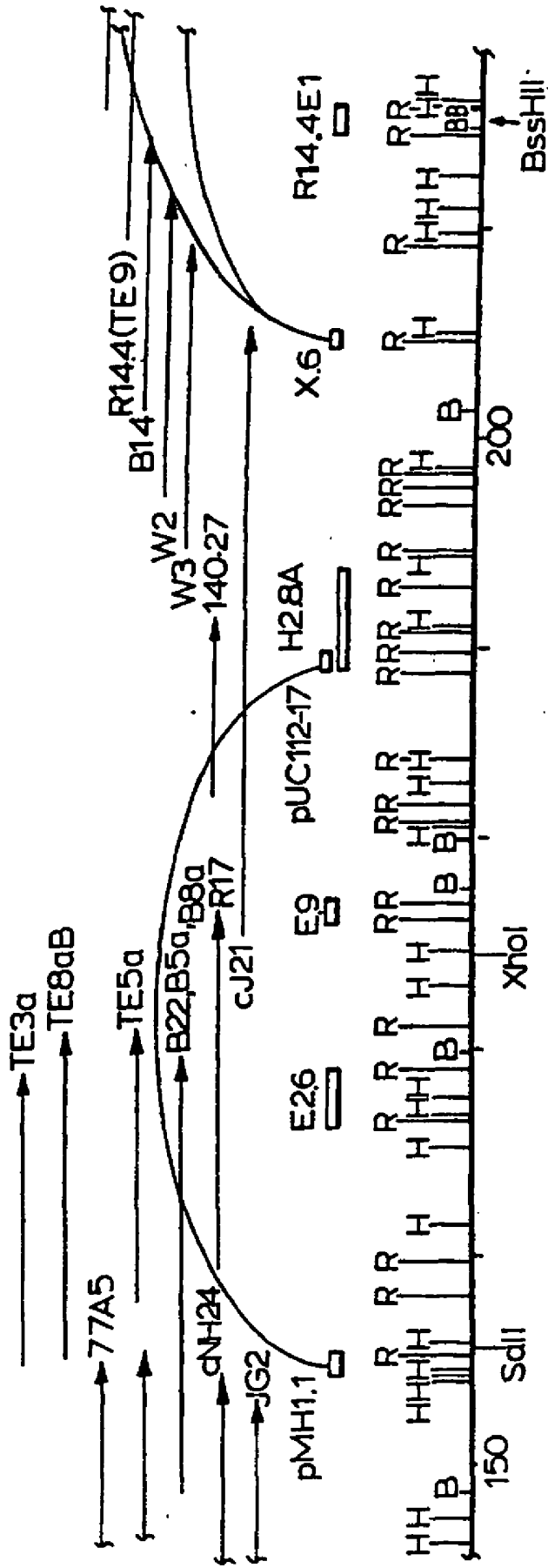


FIG.2 (cont'd)

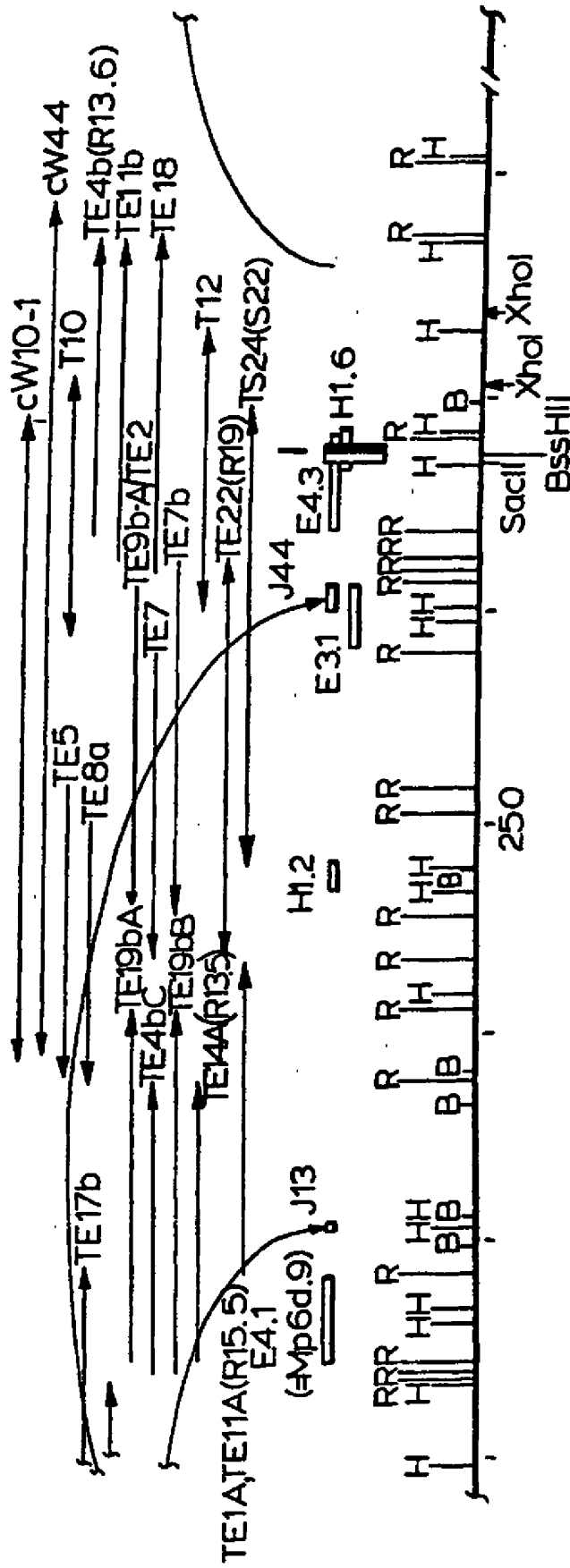


FIG.2 (cont'd)

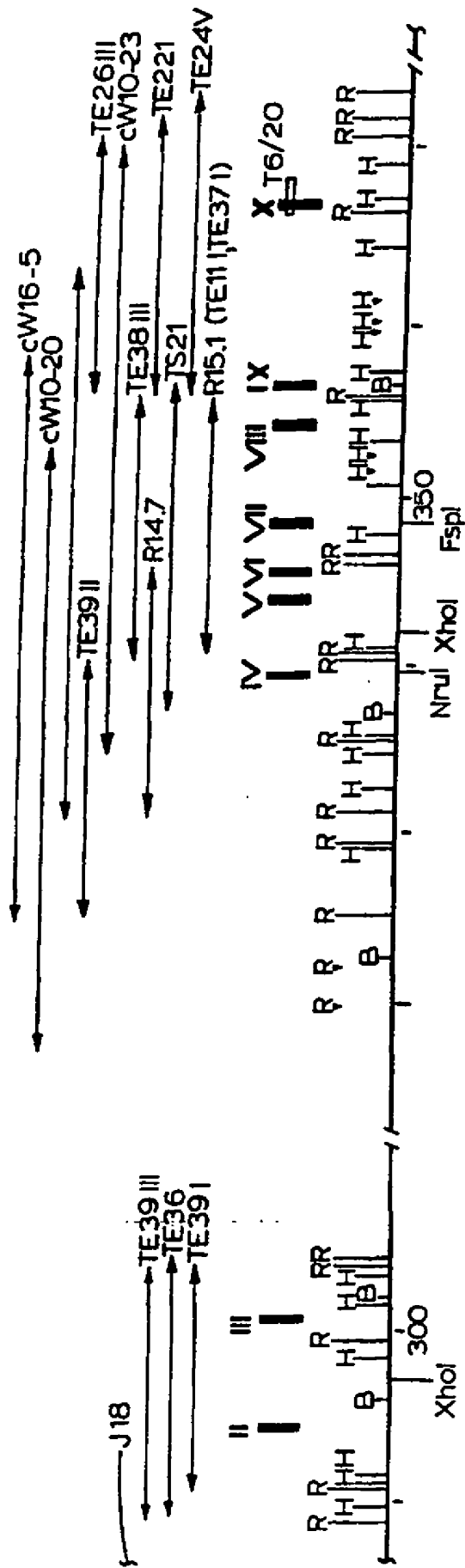


FIG.2 (cont'd)

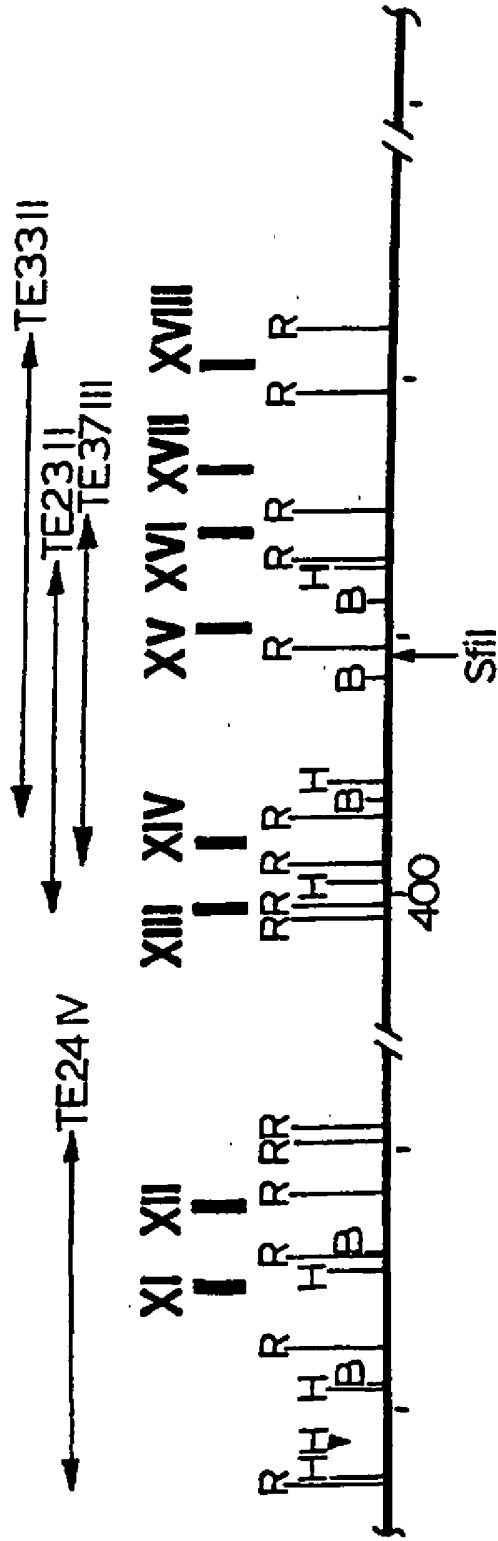


FIG.2 (cont'd)

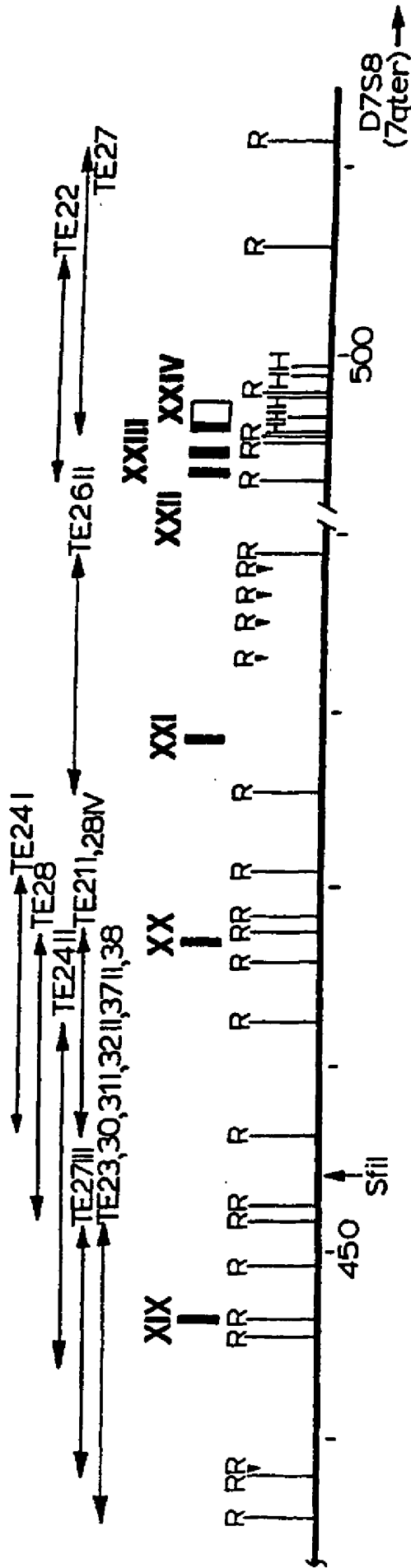


FIG. 3

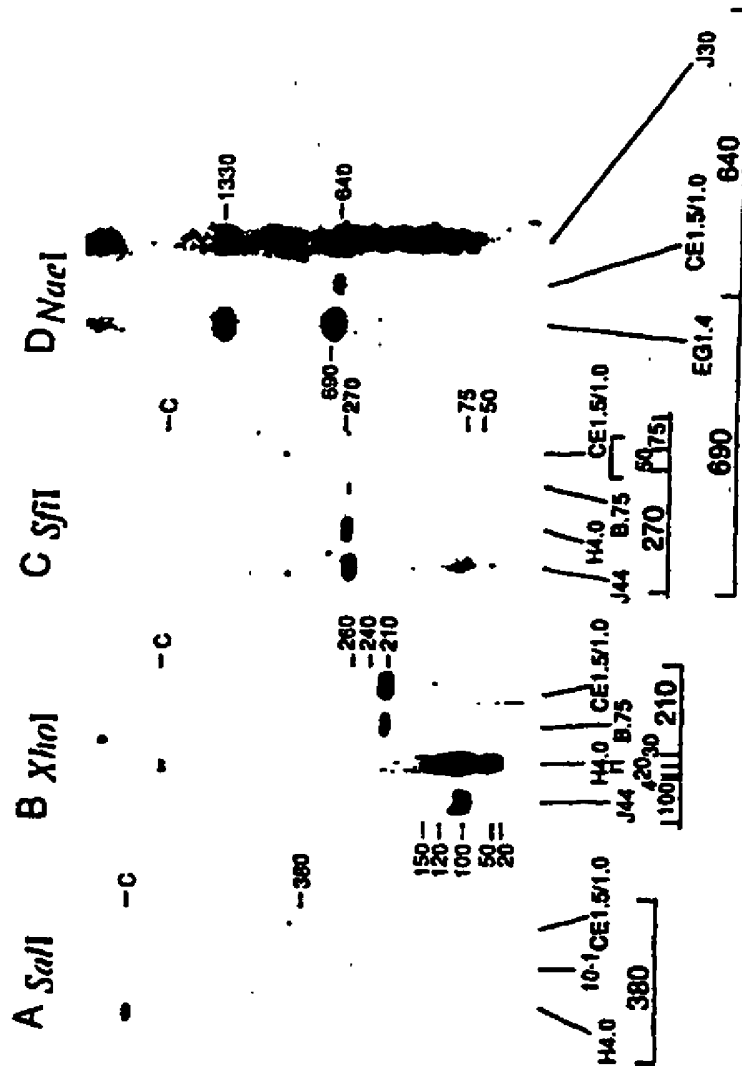


FIG.3E.

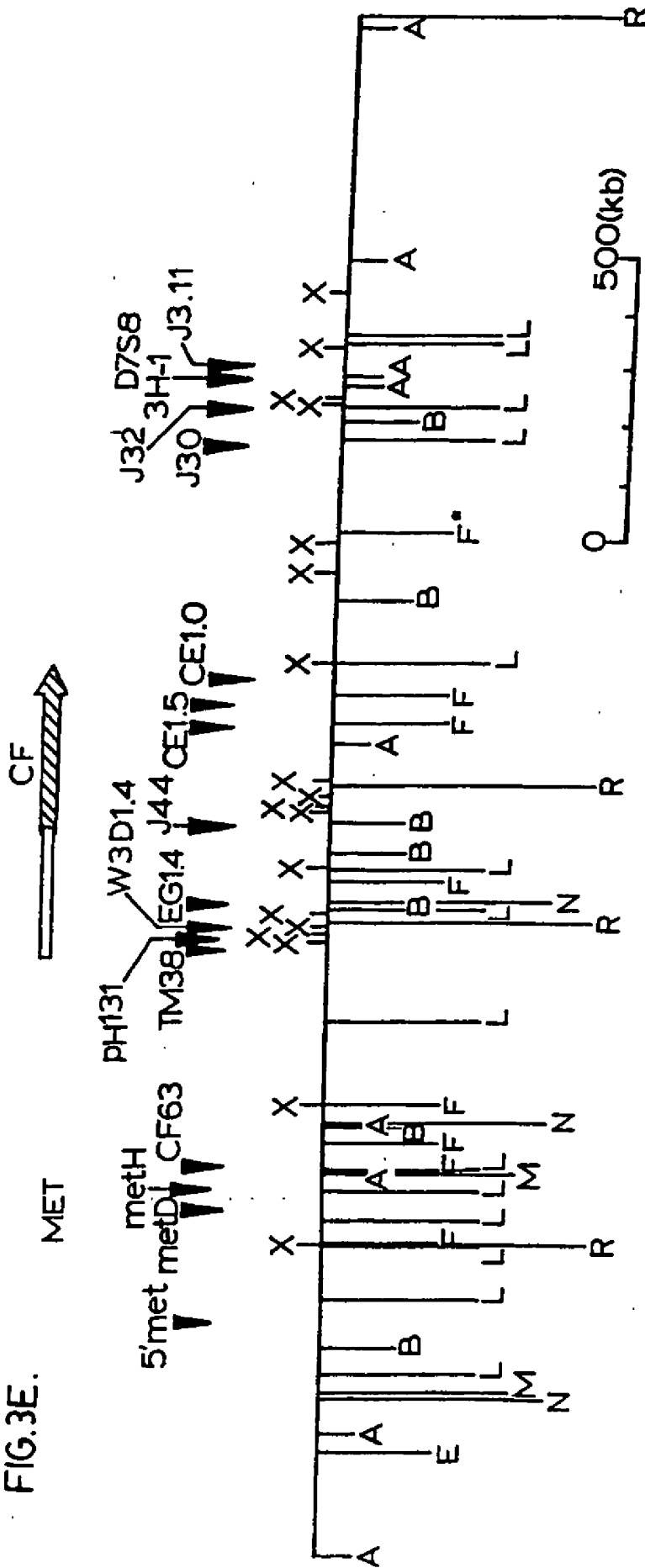




FIG. 4

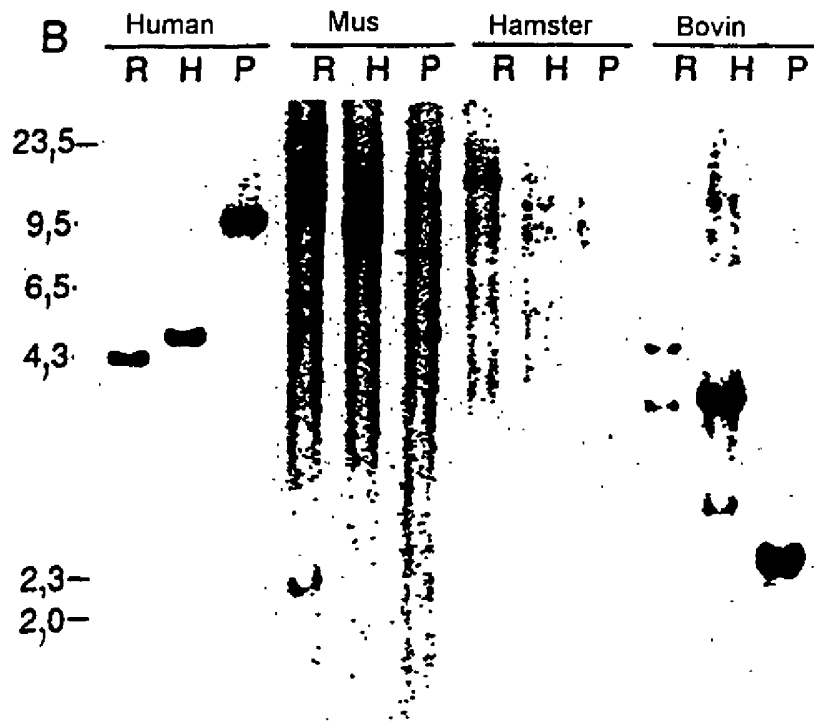
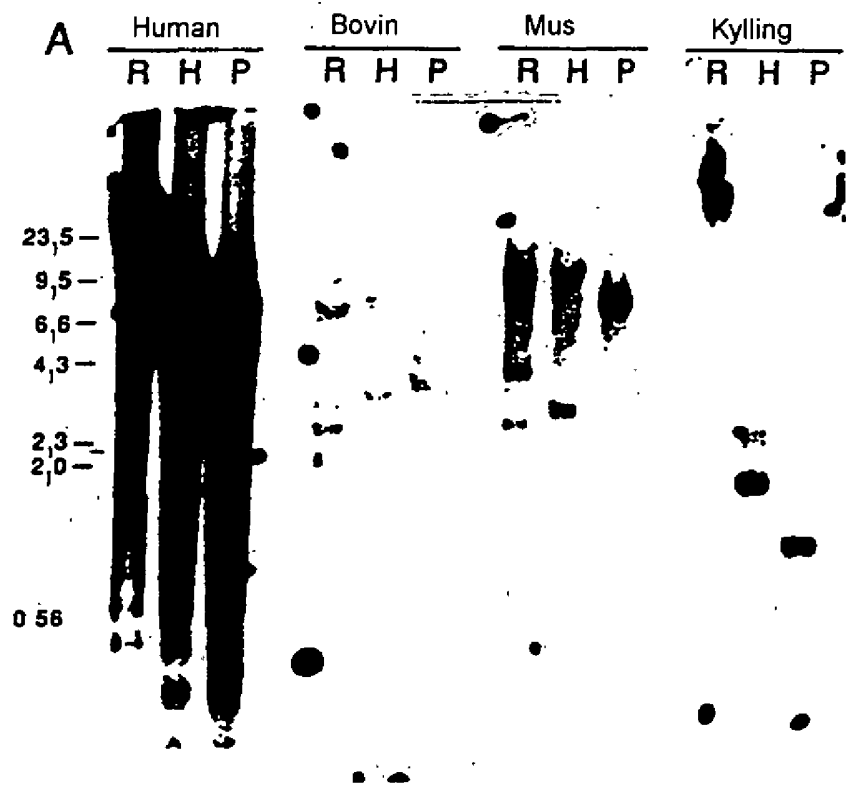
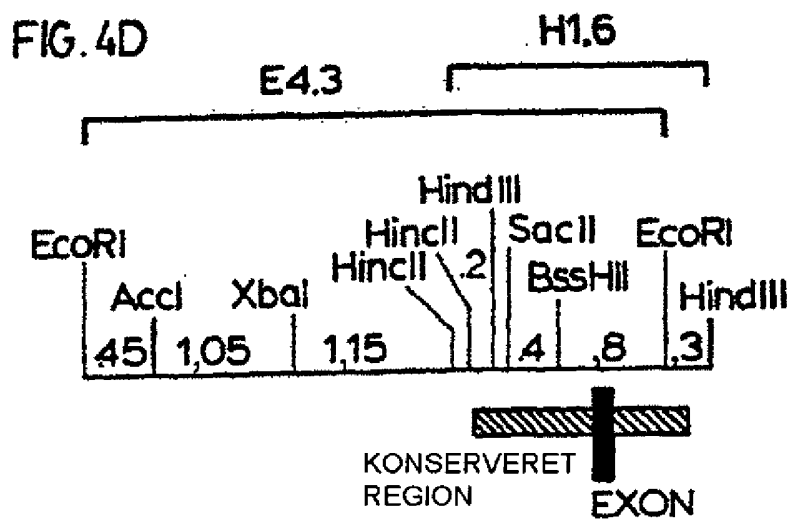
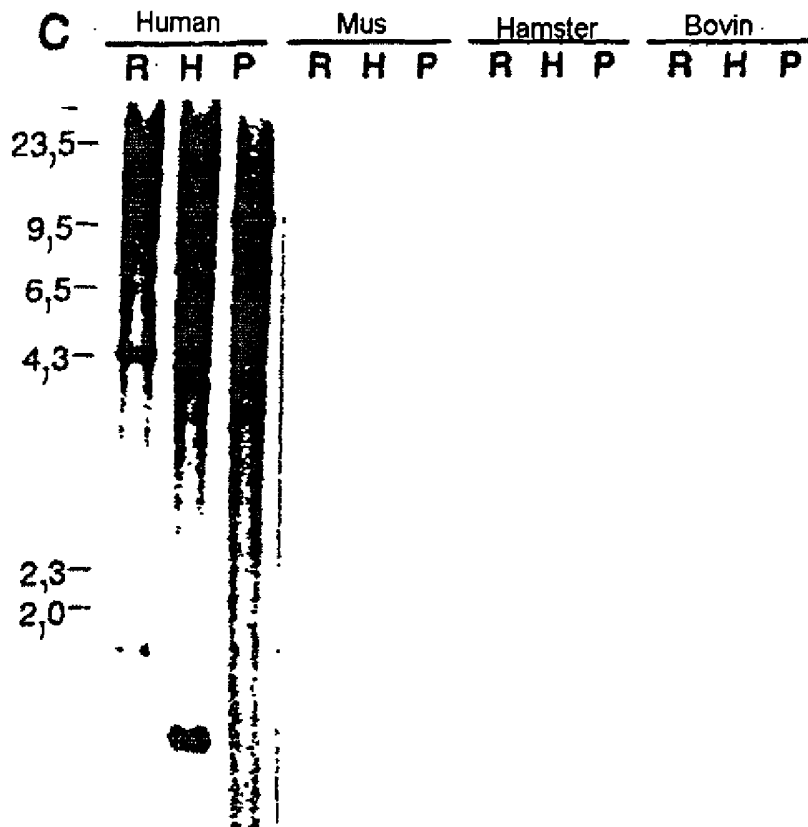


FIG. 4



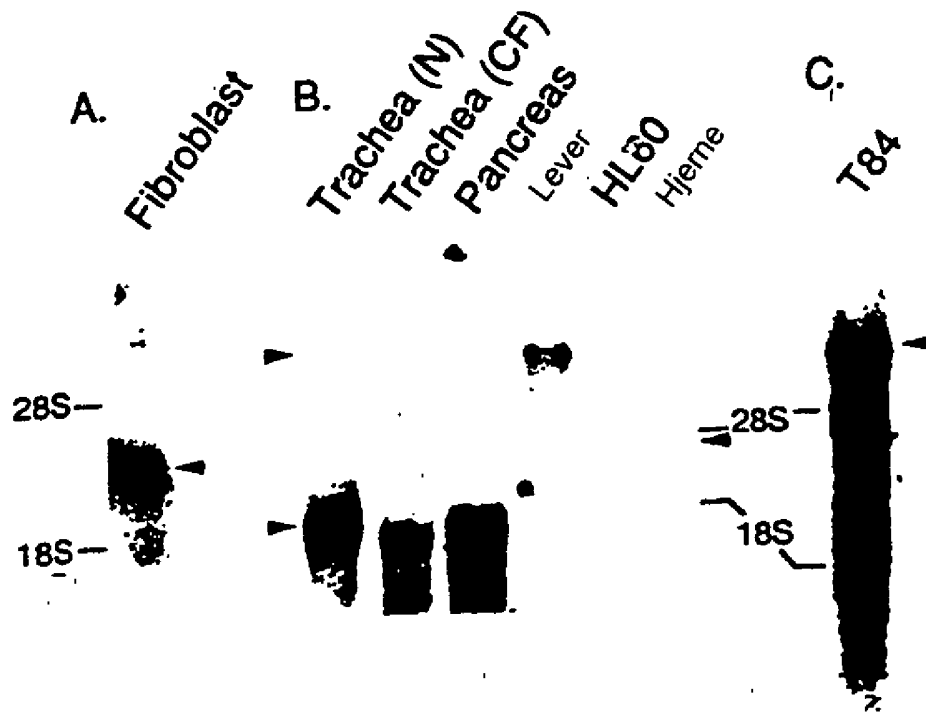


FIG. 6



FIG. 7.

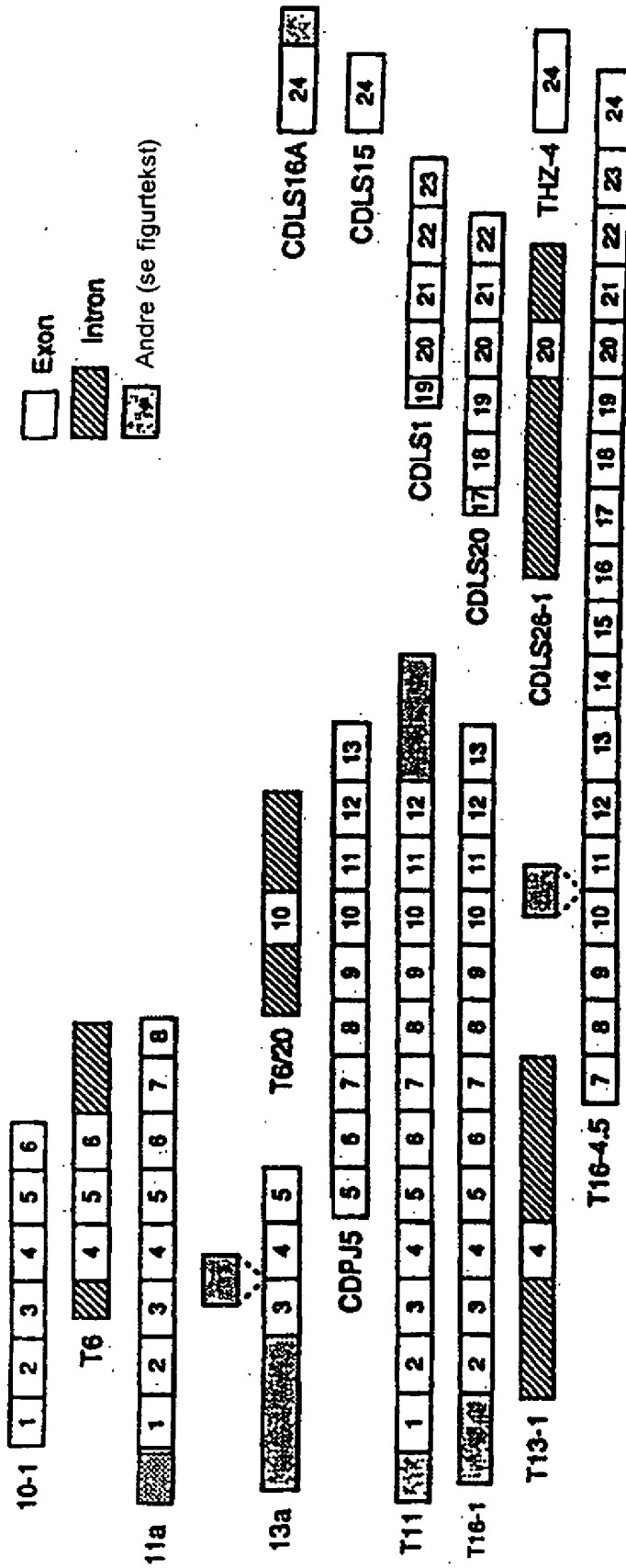
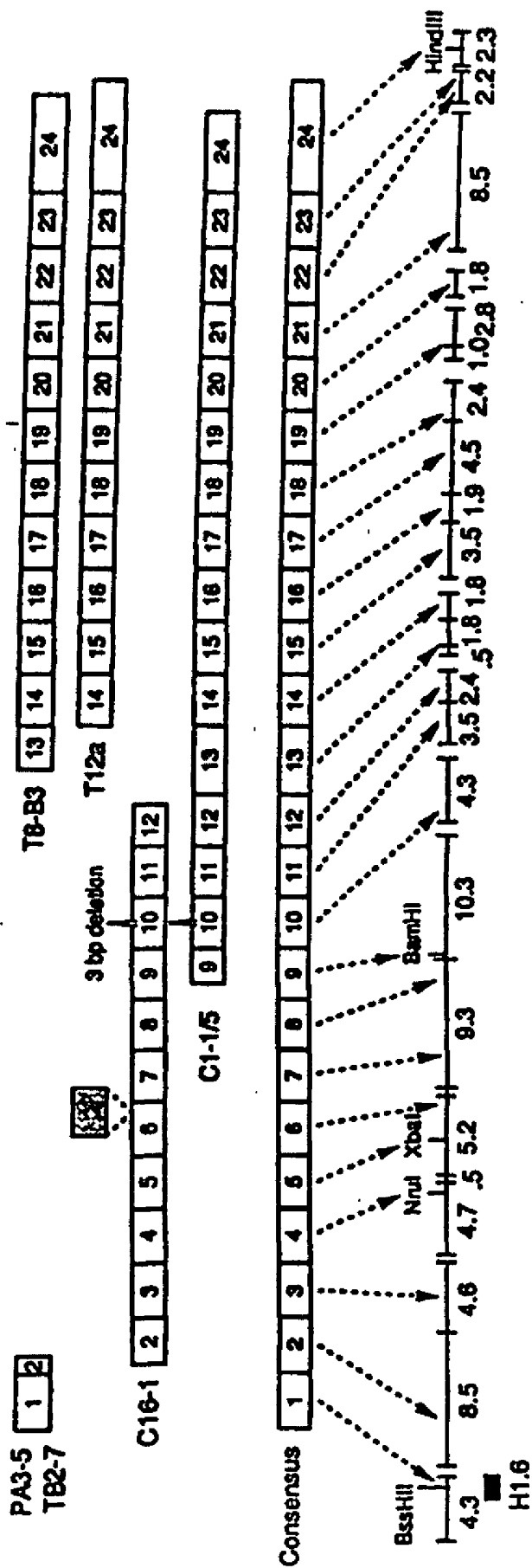


FIG 7. (cont'd)



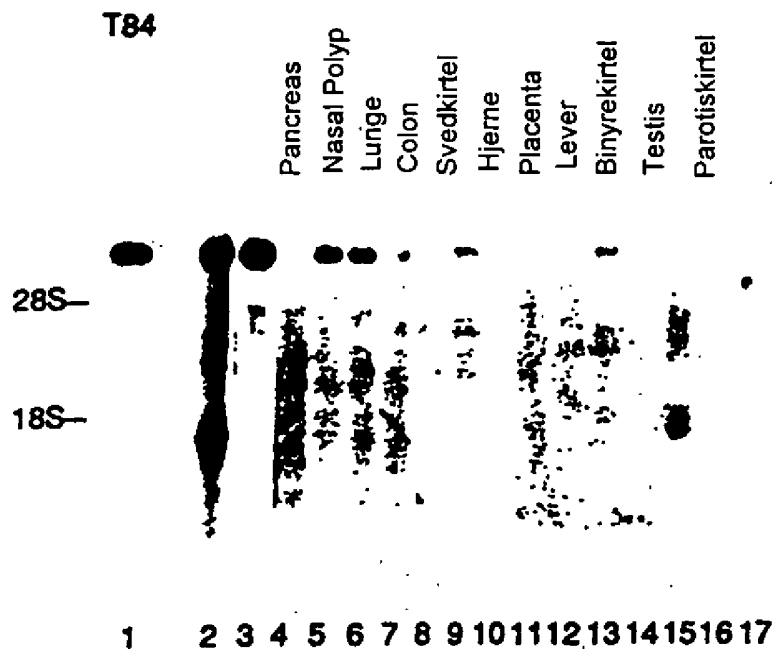


Fig. 8

FIG 9A

EcoRI  
HindIII

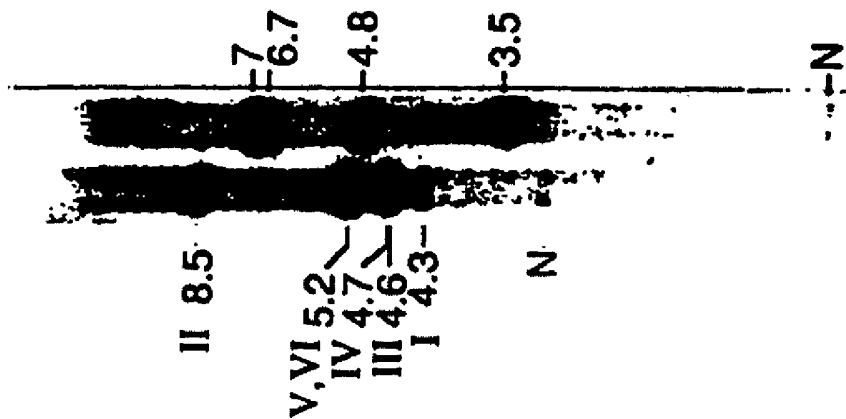


FIG 9B

EcoRI  
HindIII

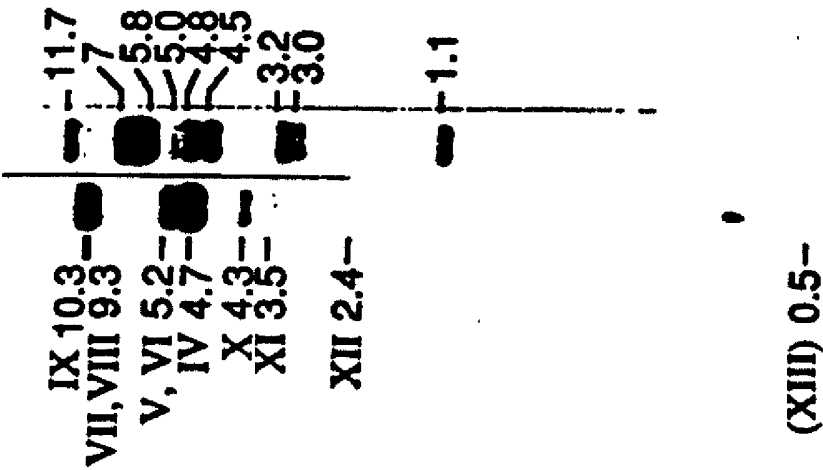


FIG 9C

EcoRI  
HindIII

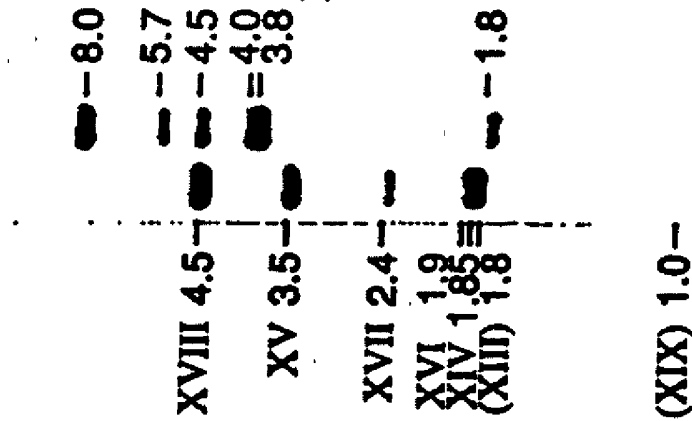


FIG 9D

EcoRI  
HindIII

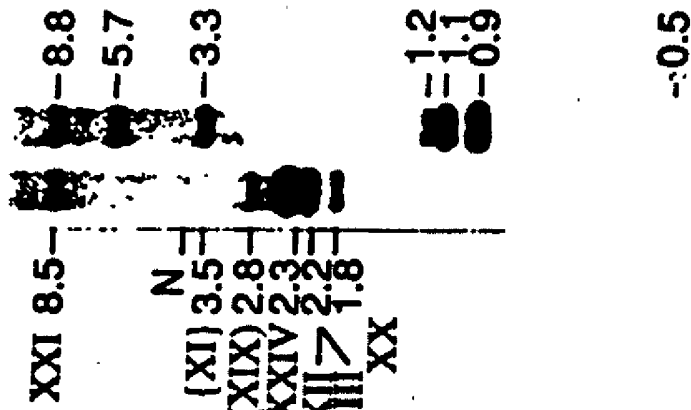


FIG. 10A

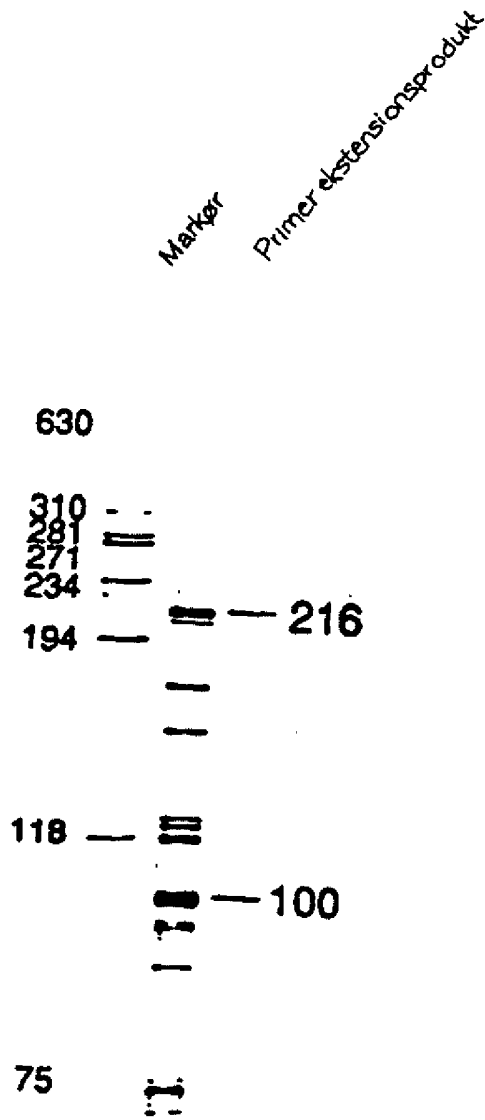




Fig. 10 B

B 1

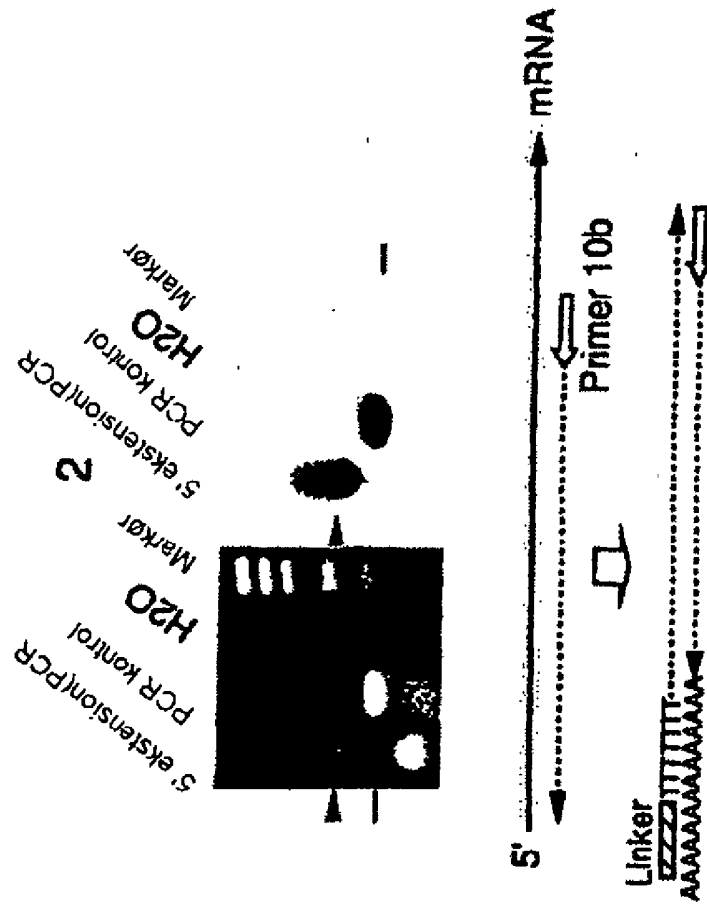


FIG. 10C

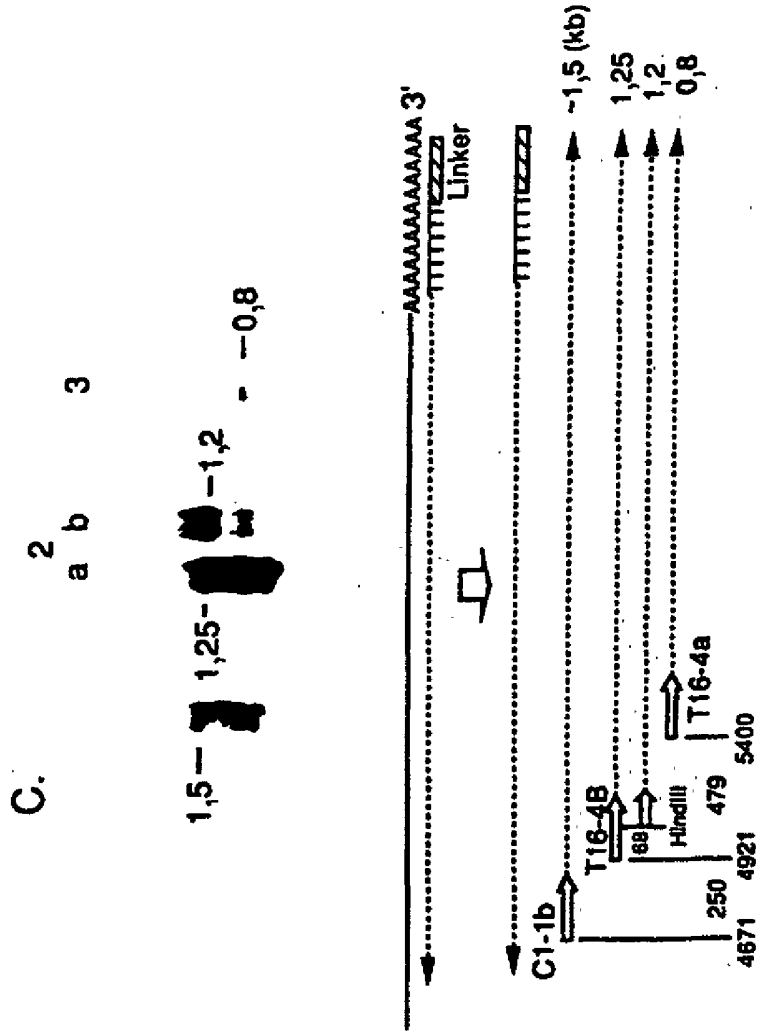


FIG. 11.

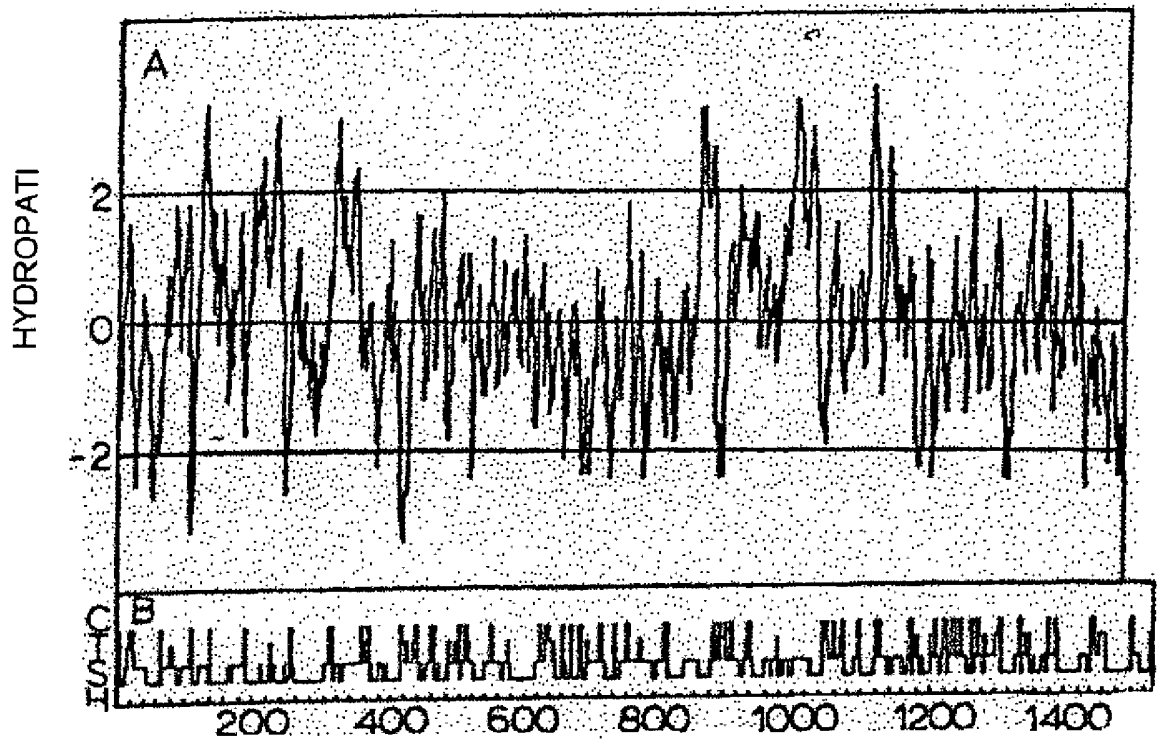


FIG. 12

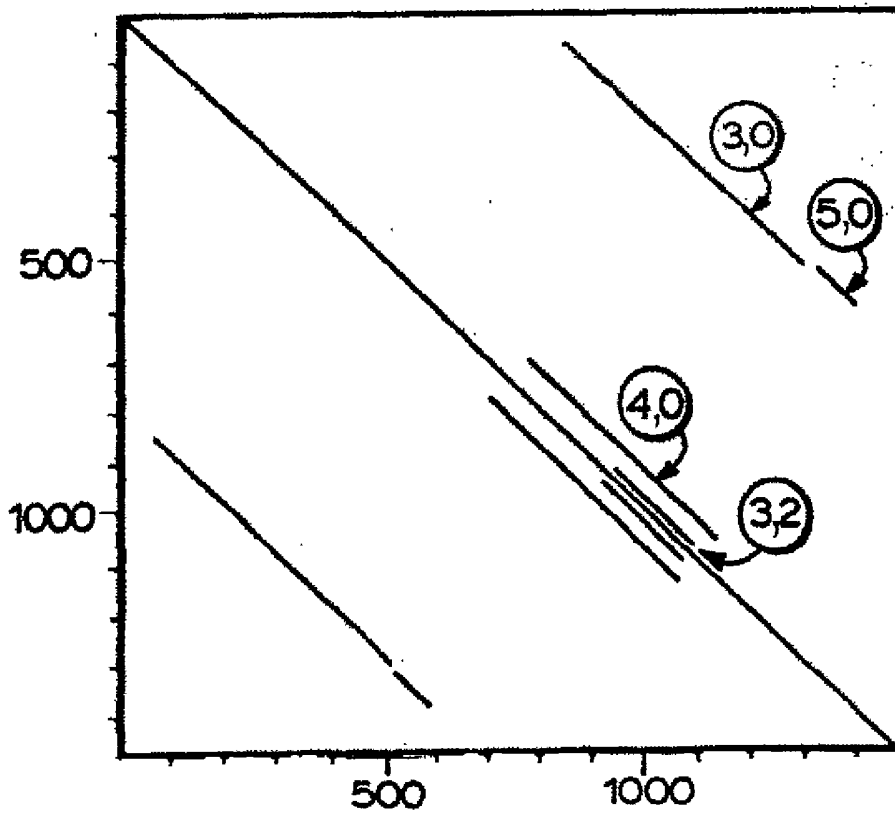


FIG. 13

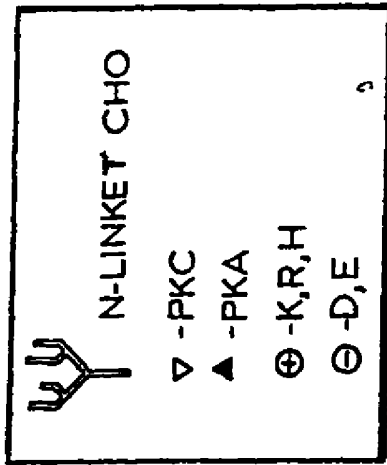
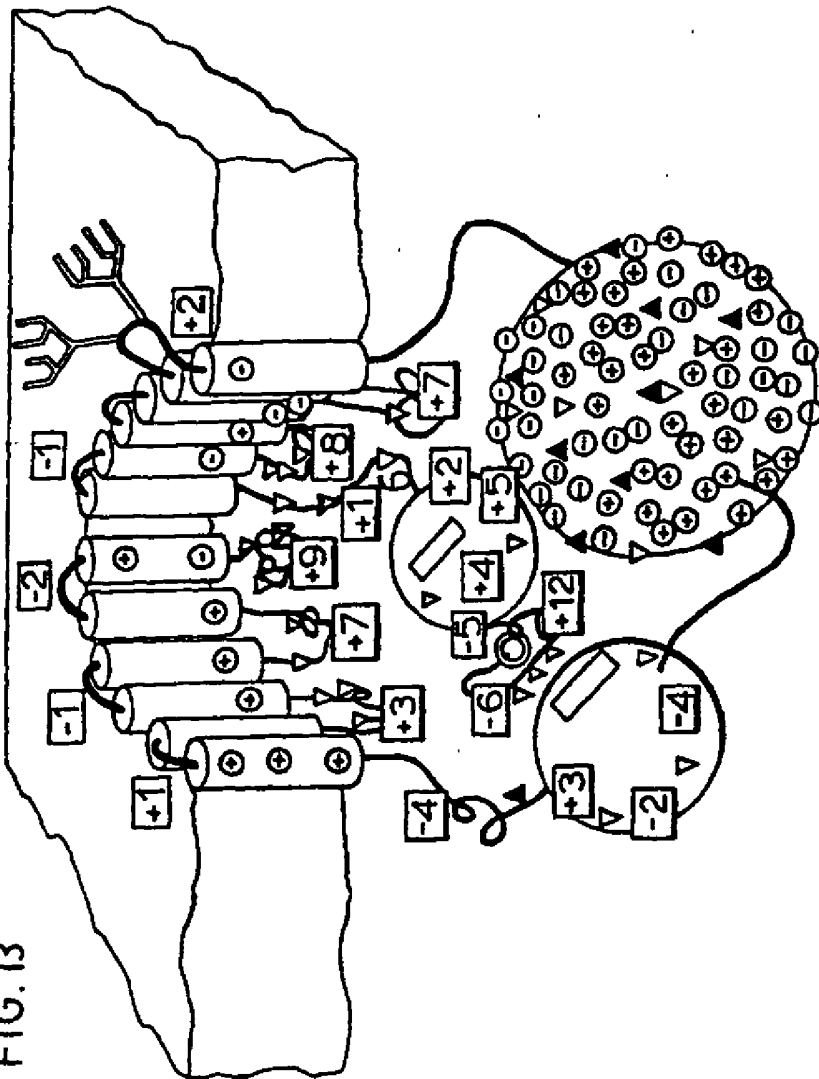


FIG.14.

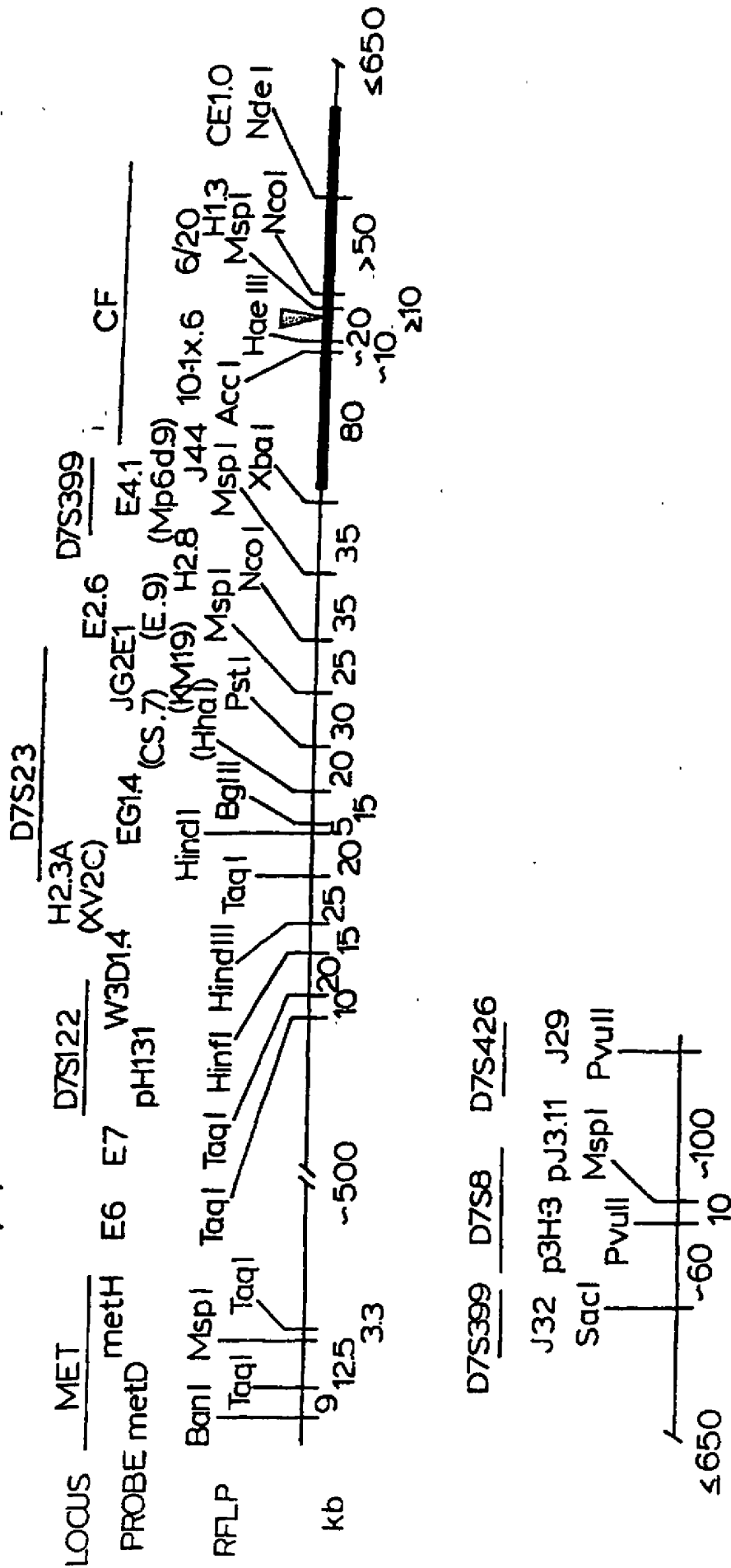
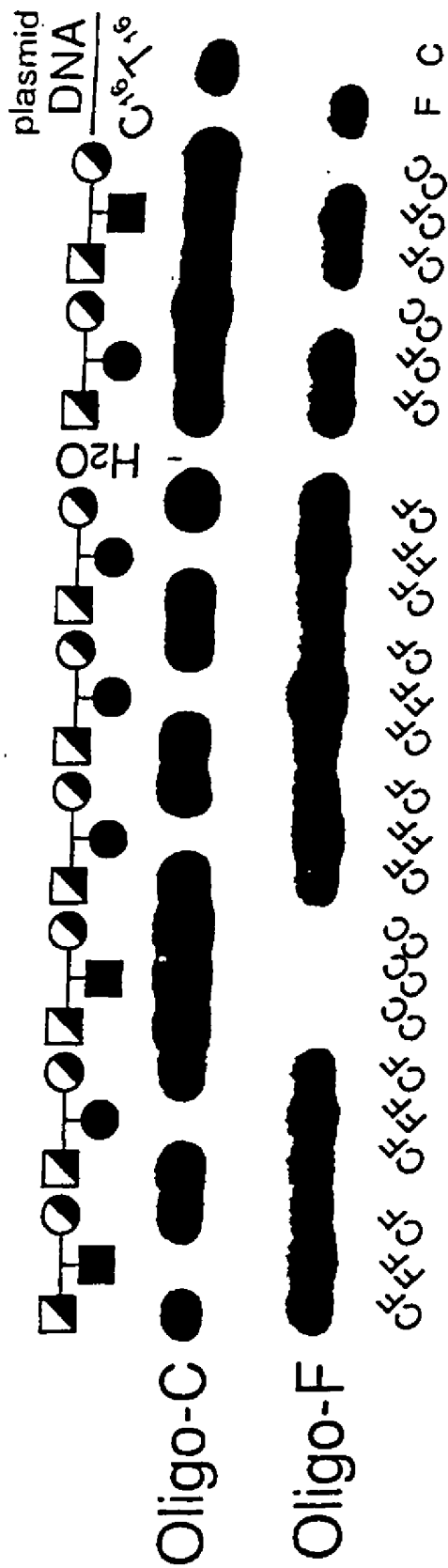


Fig. 15



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' CTTTATAGTAGAAACCAC 5'

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

FIG. 16.

ISFCSQSWIMP\*GIK-ENIIFGVSYD  
 DSITLQQRKAFGVIPQKVFIFSGTFR  
 IGWVSQEPVLFATTI-AENIRYGRENV  
 LGIVSQEPILFDCSI-AENIAYGDNSR  
 IGWVSQEPVLFATTI-AENIRYGRENV  
 LGEVSQEPILFDCSI-AENIAYGDNSR  
 IGWVSQEPVLSFTTI-AENIRYGRGNV  
 LGIVSQEPILFDCSI-AENIAYGDNSR  
 IGWVSQEPVLSFTTI-AENIAYGDNSR  
 FSIVSQEPMLFNSI-KNNIKYSLYSL  
 ITVVEQRC\*TLFNDTL-RKNILLGSTD  
 ISVVEQKPLLENGTI-RDNLTLYGLQDE  
 VGVVLDQDNVLLNRSI-IDNISLAPGMS  
 RCAYVQDDDLFIGLIAREHLIFQAMVR  
 MSFVFOHYALFKHMTVYENISFGLRLR  
 YLSQQQT\*PPFATPVWHYLLTLHQDKTR  
 VGMVFOKPT\*PPMSI-YDNIAFGVRLF  
 GIMVFOHFNLSHMTVLENVMEAPIQV  
 VGMVFOYALYPHLSVAENMSFGLKPA  
 ISMIFQDPM\*TSLNPFYMRVGEQLMEVLM  
 IQMIFQDPLASLNPRMTIGEIIAEPRLR  
 AGIIHQELNLI\*PQLTIAENIFLGREFV  
 ISEDRKRDGLVLGMSVKENMSLTALRY  
 TYTG\*VFTPVRELFAGVPE\*SRARGYTPG  
 IGIVSQEDNLDLE\*FTVRENLLVYGRYF  
 IGMIFQDHHLLMDR\*TVYD\*NV\*AI\*PL\*IA

FSLGTPVLKDI\*NF\*IERGQLLAVAGSTGAGK\*TSLLMMIMG  
 YTEGGNAILEN\*IS\*ISPCQRVGLLGR\*TS\*GK\*STLLSAFLR  
 PSRKEVKILKGLN\*LVKQSGQ\*TV\*LV\*GNS\*CG\*KS\*TV\*VQLMQR  
 PTRPDIPVLQGLS\*LEV\*KK\*Q\*TL\*LV\*GSS\*CG\*KS\*TV\*VQLLER  
 PSRSEVQILKGLN\*LVKSGQ\*TV\*LV\*GNS\*CG\*KS\*TV\*VQLMQR  
 PTRPNIPVLQGLS\*LEV\*KK\*Q\*TL\*LV\*GSS\*CG\*KS\*TV\*VQLLER  
 PSRANIKILKGLN\*LVKSGQ\*TV\*LV\*GNS\*CG\*KS\*TV\*VQLLQR  
 PTRANVPVLQGLS\*LEV\*KK\*Q\*TL\*LV\*GSS\*CG\*KS\*TV\*VQLLER  
 DTRKDV\*E\*YK\*DL\*SF\*TL\*KE\*KT\*Y\*AF\*V\*GE\*SG\*CK\*ST\*IL\*KL\*IE  
 ISRPNVP\*Y\*KN\*LS\*FT\*CD\*SK\*KT\*AI\*V\*GE\*TS\*G\*K\*ST\*FM\*NL\*LLR  
 PSRPS\*E\*AV\*LV\*KN\*V\*SN\*LF\*S\*AG\*Q\*TF\*IV\*G\*KS\*G\*K\*ST\*LS\*NL\*LLR  
 PSAPT\*AF\*V\*Y\*KN\*MF\*DM\*FC\*G\*Q\*TL\*GI\*GE\*SG\*TK\*ST\*LV\*LL\*TK  
 YKPDSP\*V\*IL\*DN\*IN\*IS\*IK\*Q\*E\*V\*IG\*IV\*GR\*SG\*G\*K\*ST\*IK\*LI\*QR  
 IPAPRKHLL\*KN\*V\*CG\*V\*AY\*P\*G\*E\*LL\*AV\*MG\*SS\*G\*AG\*KT\*TL\*LN\*AL\*AF  
 KSLGNL\*KL\*DR\*V\*SL\*Y\*VP\*K\*F\*SL\*IAL\*LG\*P\*SG\*G\*KS\*LL\*RI\*LAG  
 QDVA\*E\*STR\*LG\*P\*LS\*GE\*VR\*AG\*RI\*LL\*H\*LV\*G\*P\*NG\*AG\*K\*ST\*LI\*ARI\*AG  
 F\*Y\*Y\*G\*K\*F\*H\*AL\*KN\*IN\*LD\*T\*AK\*Q\*V\*TA\*FI\*G\*P\*SG\*G\*K\*ST\*LL\*RT\*FNK  
 RRYGG\*H\*E\*V\*LV\*KG\*V\*SL\*Q\*AR\*AG\*DV\*IS\*II\*G\*SS\*G\*K\*ST\*FL\*RC\*INF  
 KAWGE\*V\*V\*SK\*DI\*NI\*DI\*H\*E\*G\*E\*F\*V\*F\*V\*G\*P\*SG\*G\*K\*ST\*LL\*RM\*I\*AG  
 TPDGD\*V\*TA\*V\*ND\*LN\*FT\*LR\*AG\*E\*TL\*G\*IV\*G\*E\*SG\*G\*K\*ST\*AF\*AL\*MG  
 QPPK\*TL\*KA\*V\*DG\*V\*TL\*RL\*Y\*E\*G\*E\*TL\*G\*V\*G\*E\*SG\*G\*K\*ST\*FA\*RA\*II\*G  
 KAVPG\*V\*KA\*LS\*GA\*AL\*NV\*Y\*P\*GR\*V\*MA\*LV\*G\*E\*NG\*AG\*K\*ST\*MM\*K\*V\*LT\*G  
 VDNLC\*G\*P\*G\*V\*ND\*V\*SF\*TL\*RK\*GE\*LL\*G\*V\*SG\*LM\*G\*AG\*RT\*EL\*MK\*V\*LY\*G  
 LTGAR\*GN\*NL\*KD\*V\*TL\*LP\*V\*GL\*FT\*CI\*TV\*G\*SG\*G\*K\*ST\*LN\*DT\*LF  
 KSYGG\*K\*IV\*V\*ND\*LS\*FT\*IA\*AG\*E\*CF\*GL\*LP\*NG\*AG\*K\*ST\*IR\*MI\*LG  
 AYLGGR\*Q\*AL\*Q\*G\*V\*TF\*HM\*Q\*P\*G\*E\*MA\*FL\*TG\*H\*SG\*G\*K\*ST\*LL\*KL\*IGC

CFTR (N)  
 CFTR (C)  
 hmdr1 (N)  
 hmdr1 (C)  
 mmdr1 (N)  
 mmdr1 (C)  
 mmdr2 (N)  
 mmdr2 (C)  
 pfmdr (N)  
 pfmdr (C)  
 STE6 (N)  
 STE6 (C)  
 hlyB  
 white  
 MbpX  
 BtuD  
 PstB  
 hisP  
 malK  
 oppD  
 oppF  
 RbsA (N)  
 RbsA (C)  
 UvrA  
 NodI  
 FtsE

FIG. 16. (cont'd)

CFTR (N)	GEGGITLSGGQRARISLARAVYKDADLYLLDSPFGLDVLTEK
CFTR (C)	VDGGCVLSHGKQLMCLARSVLSKAKILLLLDEPSAHLDPVTYQ
hmdr1 (N)	GERGAQLSGGQKORIAARALVRNPKILLLDEATSALDTESEA
hmdr1 (C)	GDKGTLLSGGQKORIAARALVRQPHILLLDEATSALDTESEK
mmdr1 (N)	GERGAQLSGGQKORIAARALVRNPKILLLDEATSALDTESEA
mmdr1 (C)	GDKGTQLSGGQKORIAARALVRQPHILLLDEATSALDTESEK
mmdr2 (N)	GDRGAQLSGGQKORIAARALVRNPKILLLDEATSALDTESEA
mmdr2 (C)	GDKGTQLSGGQKORIAARALVRQPHILLLDEATSALDTESEK
pfmdr (N)	GSNASKLSGGQKORISIAARIMRNPKILLLDEATSSLDNKSEY
pfmdr (C)	PYGKS-JSGGQKORIAARALLREPKILLLDEATSSLDNSNSEK
STE6 (N)	GTGGVTLSGGQQQORVAIARAFIRDTPILFLDEAVSALDIVHRN
STE6 (C)	RIDTTLSSGGQAQLCIARALLRKS KILLILDECTSAALDSVSSS
hlyB	GEQAGLSGGQKORIAARALVNNPKILIFDEATSALDYASEH
White	PGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSFTA
MbpX	FEPYPAQLSGGQKORVALARSLAIQPDLLL-DEPFGALDGE
Btud	GRSTNQLSGGEWQVRVLAAVLQITLLLLDEPMNSLDVAQOSA
pstB	HQSGYSLSGGQQQORLCIARGIAIRPEVLLLDEPCSAALDPIS
hisP	GKYPVHLSGGQQQORVSIARALAMEPDLFFDEPTSAALDPELVG
malK	DRKPKALS GGQRQORVAIGRTLVAEPSVFLLEP LSNLDAALRV
oppD	KMPHEFSGGMRQORVMIAMALLCRPKLLIADEPTTALDVTVQA
oppF	NRYPHEFSGGQCORIGIARALILEPKLIICDDAVSALDVS IQA
RbsA (N)	DKLVGDLSIGDQQMVEIAKVLSESKVIIMDEPTCALIDTETE
RbsA (C)	EQAIGLLSGGNQKVAIARGLMTRPKVLIILDEPTPGVDVVGAKK
UvrA	GQSATTLSSGGEAQRVKLARELSKRGLYILDEPTTGLHFADIQQ
NodI	NTRVADLSGGMKRRLTLAGALINDPQLLILDEPTTGLDPHARH
FtsE	KNFFIQLSSGGEQQRVGIARAVVNKPAVLLADEPTGNLDDALSE



FIG. 17

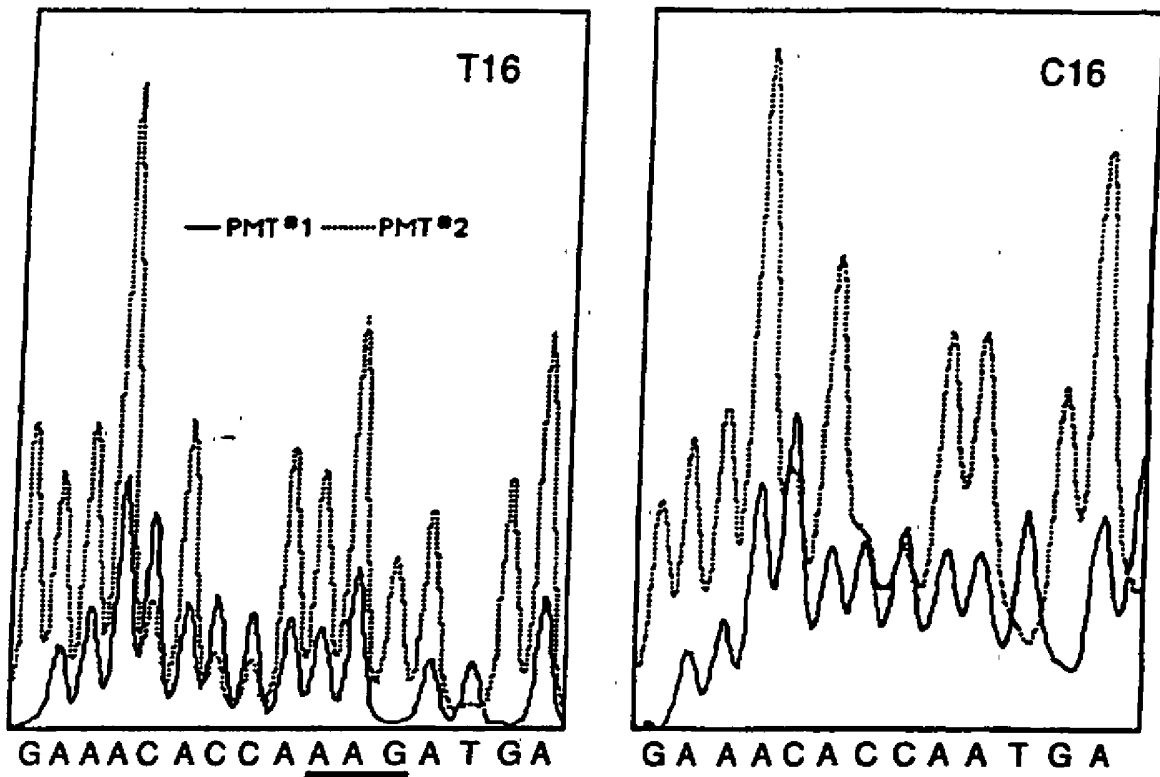


FIG. 18 A

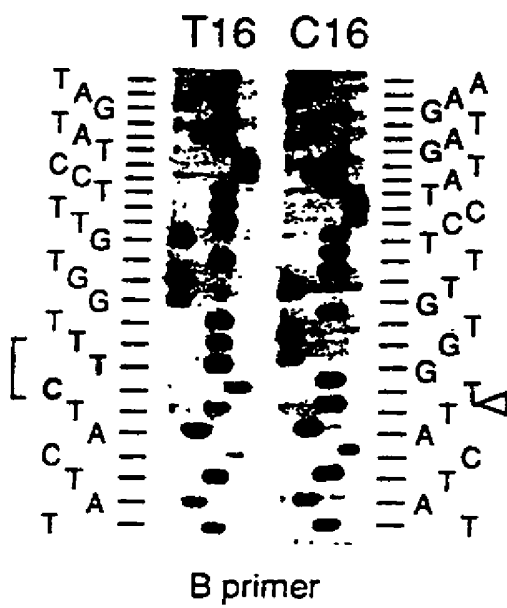


FIG. 18 B

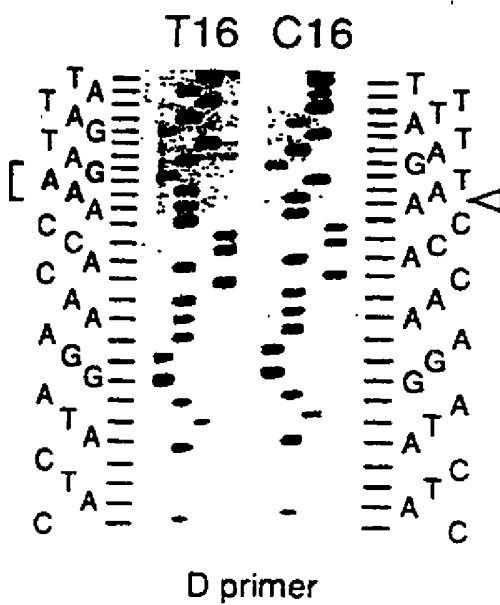


FIG. 19A

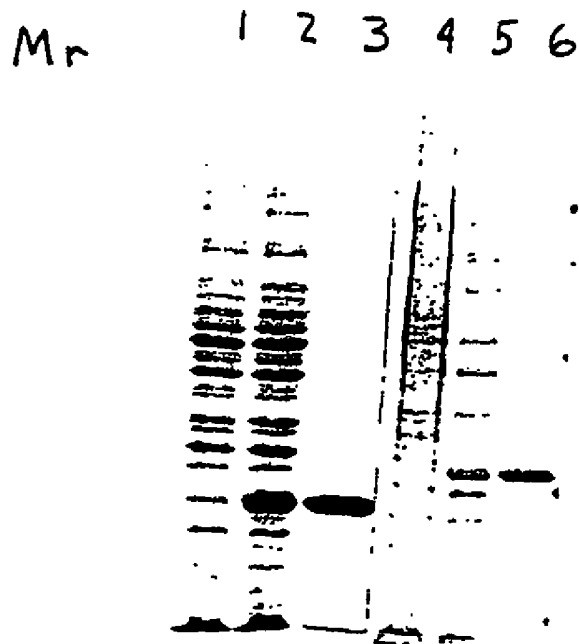


FIG. 19B

