



19 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENT- UND
MARKENAMT

12 Übersetzung der
europäischen Patentschrift

97 EP 0 667 900 B 1

10 DE 691 32 616 T 2

51 Int. Cl. 7:
C 12 N 15/12

C 12 Q 1/68
C 12 N 1/21
C 12 N 1/19
C 12 N 1/15
C 12 N 5/10
C 12 P 21/02
G 01 N 33/68
A 01 K 67/02
G 01 N 33/577
G 01 N 33/534

DE 691 32 616 T 2

- 21 Deutsches Aktenzeichen: 691 32 616.9
- 86 PCT-Aktenzeichen: PCT/CA91/00009
- 96 Europäisches Aktenzeichen: 91 902 210.3
- 87 PCT-Veröffentlichungs-Nr.: WO 91/10734
- 86 PCT-Anmeldetag: 11. 1. 1991
- 87 Veröffentlichungstag der PCT-Anmeldung: 25. 7. 1991
- 97 Erstveröffentlichung durch das EPA: 23. 8. 1995
- 97 Veröffentlichungstag der Patenterteilung beim EPA: 23. 5. 2001
- 47 Veröffentlichungstag im Patentblatt: 21. 2. 2002

30 Unionspriorität:

2007699	12. 01. 1990	CA
2011253	01. 03. 1990	CA
2020817	10. 07. 1990	CA

72 Erfinder:

TSUI, Lap-Chee, Toronto, CA; ROMMENS, M.,
Johanna, Willowdale, CA; KEREM, Bat-sheva, 91
904 Jerusalem, IL

73 Patentinhaber:
HSC Research Development Corp., Toronto,
Ontario, CA

74 Vertreter:
Kuhnen & Wacker Patentanwaltsgesellschaft mbH,
85354 Freising

84 Benannte Vertragsstaaten:
CH, DE, FR, GB, IT, LI, NL, SE

54 INTRONS UND EXONS DES GENS FÜR CYSTISCHE FIBROSE SOWIE MUTATIONEN AN MEHREREN STELLEN DES GENS

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

DE 691 32 616 T 2



Die Erfindung betrifft allgemein das cystische Fibrose (CF)-Gen und genauer die Identifikation, die Isolierung und das Klonieren der DNA-Sequenz, die mit Mutanten
5 des CF-Gens korrespondiert, ebenso wie ihre Transkripte, Genprodukte und genetischen Informationen an den Exon/Intron-Grenzen. Die Erfindung betrifft auch Verfahren für das Screening auf der Suche nach CF-Trägern und für den Nachweis von CF-Trägern, für das prenatale CF-Screening und für die prenatale CF-Diagnose und eine Gen-
10 therapie, die sich Rekombinationstechniken zunutze macht, und eine Arzneimitteltherapie, die Informationen nutzt, die über die DNA, das Protein und die metabolische Funktion des Proteins zur Verfügung stehen.

HINTERGRUND DER ERFINDUNG

15 Die cystische Fibrose (CF) stellt die am meisten verbreitete schwere, autosomal rezessive, genetische Störung in der weißen Bevölkerung dar. Sie betrifft in Nordamerika ungefähr 1 von 2000 Lebendgeburten [Boat et al., *The Metabolic Basis of Inherited Disease*, 6. Aufl., S. 2649 – 2680, McGraw Hill, NY (1989)]. Ungefähr 1 von
20 Personen ist Träger der Krankheit.

Obwohl die Krankheit das erste Mal in den späten 30er Jahren des 20. Jahrhunderts beschrieben wurde, ist ihr Grunddefekt immer noch unbekannt. Die wichtigsten Symptome der cystischen Fibrose sind unter anderem chronische Atemwegserkrankungen, Bauchspeicheldrüsen-Unterfunktion und erhöhte Elektrolytwerte im
25 Schweiß. Die Symptome lassen darauf schließen, daß die cystische Fibrose eine exokrine Funktionsstörung ist. Obwohl in letzter Zeit Fortschritte bei der Analyse des Ionentransports durch die Apikalmembran des Zellepithels von CF-Patienten gemacht wurden, ist nicht geklärt, ob die abnormale Regulation der Chloridkanäle den Hauptdefekt bei dieser Krankheit darstellt. Angesichts einer fehlenden Erklärung des
30 Molekularmechanismus der Krankheit wurde ein alternativer Weg eingeschlagen, um zu



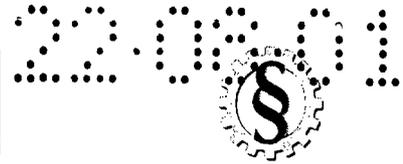
versuchen, durch direktes Klonieren des verantwortlichen Gens die Natur des Molekulardefekts aufgrund seiner Position im Chromosom zu verstehen.

Es existiert jedoch kein klarer Phänotyp, der auf diesem Weg zur exakten Natur
5 der genetischen Grundlage der Krankheit führt, oder der eine Identifikation des
cystischen Fibrose-Gens erlaubt. Die Natur des CF-Defekts war in Beziehung zu den
genetischen Daten der Bevölkerung bisher nicht leicht zu verstehen. Sowohl die
Verbreitung der Krankheit als auch ihre klinische Heterogenität wurden mit mehreren
verschiedenen Mechanismen erklärt: einer hohen Mutationsrate, einem heterozygoten
10 Vorteil, der genetischen Drift, mehrfachen Loci und der reproduktiven Kompensation.

Viele dieser Hypothesen können aufgrund der fehlenden Wissens um den Grund-
defekt nicht überprüft werden. Darum konzentrierte man sich auf alternativen Wegen
zur Bestimmung und Beschreibung des CF-Gens auf den Versuch, die Lage des Gens
15 durch Genanalyse zu identifizieren.

Ein Versuch, bei dem die Kopplung des CF-Gens an Antigen- und Proteinmarker
analysiert wurde, wurde in den 50er Jahren des 20. Jahrhunderts unternommen, aber es
wurden keine positiven Ergebnisse erzielt [Steinberg et al., *Am. J. Hum. Genet.* 8: 162
20 bis 176, (1956); Steinberg und Morton, *Am. J. Hum. Genet.* 8: 177 – 189, (1956);
Goodchild et al. *J. Med. Genet.* 7: 417 – 419, (1976)].

In jüngerer Zeit wurde es möglich, RFLPs zu verwenden, um die Kopplungs-
analyse zu erleichtern. Die erste Kopplung eines RFLP-Markers an das CF-Gen wurde
25 1985 offenbart [Tsui et al., *Science* 230: 1054-1057, 1985], wobei eine Kopplung
zwischen dem CF-Gen und dem nicht charakteristischen Marker D0CRI-917 gefunden
wurde. Der Zusammenhang wurde bei der Analyse von 39 Familien mit CF-kranken
Kindern entdeckt. Obwohl die chromosomale Lage nicht ermittelt worden war, zeigt
dies, daß die Lage des krankmachenden Gens auf etwa 1 % des menschlichen Genoms
30 oder auf etwa 30 Millionen Nucleotid-Basenpaare eingegrenzt worden war.



Die chromosomale Lage der D0CRI-917-Sonde wurde unter Verwendung von Nagetier/Mensch-Hybridzelllinien ermittelt, die verschiedene menschliche Chromosomen-Komplemente enthielten. Es wurde gezeigt, daß DOCR1-917 (und damit das CF-Gen) auf dem menschlichen Chromosom 7 liegt.

5

Es wurden weitere physische und genetische Kopplungsstudien durchgeführt, um zu versuchen, die Position des CF-Gens genau einzugrenzen. Zengerling et al. [*Am. J. Hum. Genet.* 40: 228-236 (1987)] beschreiben die Verwendung von Mensch/Maus-Somazellhybriden, um eine detaillierte physische Beziehung zwischen dem CF-Gen und den Markern, von denen man weiß, daß sie an dieses gekoppelt sind, zu erhalten. Diese Veröffentlichung zeigt, daß das CF-Gen entweder der distalen Region der q22 Bande oder der proximalen Region der q31 Bande auf dem Chromosom 7 zugeordnet werden kann.

10

Rommens et al. [*Am. J. Hum. Genet.* 43: 645-663 (1988)] liefern eine detaillierte Erörterung der Isolierung von vielen neuen 7q31 Sonden. Der dargestellte Weg führte zur Isolierung von zwei neuen Sonden, D7S122 und D7S340, die sich nahe aneinander befinden. Die Pulsfeld-Gelelektrophoresekartierung zeigt an, daß diese beiden RFLP-Marker sich zwischen zwei Markern, MET [White, R., Woodward S., Leppert M. et al., *Nature* 318: 382-384, (1985)] und D7S8 [Wainwright, B.J., Scambler, P.J. und J. Schmidtke, *Nature* 318: 384-385 (1985)], von denen man weiß, daß sie das CF-Gen einrahmen, und daher in der CF-Genregion befinden. Die Entdeckung dieser Marker liefert einen Ausgangspunkt für das Chromosomen-Walking und -Jumping.

20

Estivill et al. [*Nature* 326: 840-845 (1987)] offenbaren, daß ein Kandidaten-cDNA-Gen lokalisiert und teilweise beschrieben wurde. Sie lehren jedoch nicht die korrekte Lage des CF-Gens. Die Literaturstelle offenbart ein Kandidaten-cDNA-Gen flußabwärts von einer CpG-Insel, wobei es sich um untermethylierte, GC-Nukleotidreiche Regionen flußaufwärts von vielen Wirbeltiergenen handelt. Die chromosomale Lokalisierung der Lage des Kandidatengens wird als die XV2C-Region identifiziert.

25

30

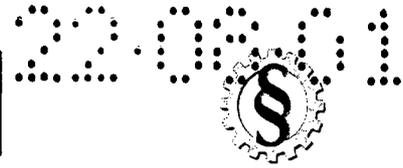


Eine Hauptschwierigkeit bei der Identifikation des CF-Gens war das Fehlen von cytologisch bestimmbar Chromosomen-Umgruppierungen oder -Deletionen, die sämtliche bisherigen Erfolge beim Klonieren von krankmachenden menschlichen Genen wegen des Wissens um ihre Kartenposition sehr erleichtert hatten.

5

Solche Umgruppierungen und Deletionen konnten cytologisch untersucht werden und infolgedessen konnte ein Zusammenhang zwischen der physischen Lage auf einem bestimmten Chromosom und der jeweiligen Krankheit hergestellt werden. Außerdem konnte aufgrund der bekannten Beziehung zwischen offen zugänglichen DNA-Sonden und cytologisch erkennbaren Veränderungen in den Chromosomen ein Zusammenhang zwischen dieser cytologischen Lage und einer molekularen Lage hergestellt werden. Das Wissen um die molekulare Lage des Gens für eine bestimmte Krankheit ermöglicht das Klonieren und Sequenzieren dieses Gens anhand von Routineverfahren, insbesondere wenn das Genprodukt bekannt ist und der Erfolg des Klonierens durch einen Immunoassay der Expressionsprodukte der klonierten Gene bestätigt werden kann.

Im Gegensatz dazu war im Stand der Technik weder die cytologische Lage noch das Genprodukt des Gens für cystische Fibrose bekannt. Mit der kürzlich gelungenen Identifizierung von MET und D7S8, Markern, die das CF-Gen einrahmten, aber dessen molekulare Lage nicht genau eingrenzten, erarbeiteten die Erfinder verschiedene neuartige Genklonierungsstrategien gemäß der Erfindung, um sich dem CF-Gen zu nähern. Die Verfahren, die bei diesen Strategien angewandt wurden, schließen das Chromosomen-Jumping von den flankierenden Markern, das Klonieren von DNA-Fragmenten aus einer definierten physischen Region unter Verwendung der Pulsfeld-Gelelektrophorese, eine Kombination aus Somazellhybrid- und molekularen Klonierungstechniken, die darauf ausgelegt waren, DNA-Fragmente aus untermethylierten CpG-Inseln in der Nähe des CF zu isolieren, die Chromosomen-Mikrodissektion und -klonierung und das Sättigungsklonieren einer großen Zahl von DNA-Markern aus der Region 7q31 ein. Anhand dieser neuartigen Strategien waren die Erfinder in der Lage, das Gen, das für die cystische Fibrose verantwortlich ist, dort zu identifizieren, wo der Stand der Technik unsicher oder, in einem Fall, sogar falsch war.



Die Anwendung dieser genetischen und molekularen Klonierungsstrategien ermöglichte die Isolierung und das cDNA-Klonieren des cystische Fibrosegens auf der Grundlage seiner Lage im Chromosom, ohne daß Genom-Umgruppierungen dabei
5 behilflich gewesen wären. Die Identifikation der normalen und mutierten Formen des CF-Gens und seiner Genprodukte ermöglichte die Entwicklung von Screening- und diagnostischen Tests auf CF unter Verwendung von Nukleinsäuresonden und Antikörpern für das Genprodukt. Durch Interaktion mit dem Defekt-Genprodukt und dem Stoffwechselweg, den dieses Genprodukt nimmt, wurde nun eine Therapie durch
10 Ergänzung mit normalem Genprodukt und durch Genmanipulation und -einbau möglich.

Das Gen, das am cystischen Fibrose-Krankheitsverlauf beteiligt ist, im folgenden „CF-Gen“ genannt, und seine funktionellen Äquivalente, wurde identifiziert, isoliert
15 und cDNA-kloniert, und seine Transkripte und Genprodukte wurden identifiziert und sequenziert. Es wurde nachgewiesen, daß eine Deletion von drei Basenpaaren, die zum Fehlen eines Phenylalaninrests im Genprodukt führt, mit den Mutationen des CF-Gens bei ungefähr 70 % der an CF erkrankten Patienten übereinstimmte, wobei in den meisten, wenn nicht allen der übrigen Fälle andere Mutationen beteiligt waren.

20 Spezifische Aspekte werden von den Erfindern in Beiträgen für Fachzeitschriften erörtert, nämlich *Science* (1989) 245 Nr. 4922, S. 1066-1073, und *Science* (1989) 245 Nr. 4922, S. 1073-1080. Diese Fachzeitschriftenartikel erörtern die CFTR-Gensequenz und die vorstehend genannte Mutation.

25

ZUSAMMENFASSUNG DER ERFINDUNG

Gemäß der Erfindung wurden auch andere Basenpaar-Deletionen oder -Abänderungen, die zum Fehlen von Aminosäureresten im Genprodukt führten, nachgewiesen. Gemäß der Erfindung wurden auch andere Nucleotid-Deletionen oder -Ab-
30



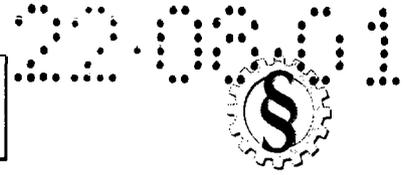
änderungen, die zu Mutationen der DNA-Sequenz führen, die Frameshift- oder Spleiß-Mutationen zur Folge haben, bestimmt.

Mit der Identifikation und Sequenzierung des Mutantengens und seines Gen-
5 produkts können Nukleinsäuresonden und Antikörper, die sich auf das Mutantengen-
produkt richten, in einer Reihe von Hybridisations- und immunologischen Assays ver-
wendet werden, um das CF-Defektgen oder -genprodukt aufzuspüren oder nach-
zuweisen. Assaykits für ein solches Screening und eine solche Diagnose können eben-
falls bereitgestellt werden. Die genetische Information, die aus den Intron/Exon-
10 Grenzen stammt, ist bei verschiedenen Screening- und Diagnoseverfahren ebenfalls sehr
nützlich.

Die Patiententherapie durch die Supplementation mit dem normalen Genprodukt,
dessen Produktion mittels genetischer und rekombinanter Techniken amplifiziert
15 werden kann, oder seinem funktionellen Äquivalent, ist nun ebenfalls möglich. Die
Korrektur oder Modifikation des Defektgenprodukts ist nun auf dem Wege der Arznei-
mitteltherapie möglich. Darüber hinaus kann die cystische Fibrose durch Gentherapie
geheilt oder kontrolliert werden, indem man den Gendefekt *in situ* oder mittels re-
kombinanter oder anderer Vehikel, die eine DNA-Sequenz liefern, die zur Expression
20 des normalen Genprodukts für die Zellen des Patienten in der Lage ist, korrigiert.

Gemäß einem weiteren Aspekt der Erfindung umfaßt ein gereinigtes CF-
Mutantengen eine DNA-Sequenz, die für eine Aminosäuresequenz für ein Protein
kodiert, wobei das Protein, wenn es in Zellen des menschlichen Körpers exprimiert
25 wird, mit einer geänderten Zellfunktion assoziiert wird, die mit der genetischen Krank-
heit cystische Fibrose korreliert.

Gemäß einem weiteren Aspekt der Erfindung umfaßt ein gereinigtes RNA-
Molekül eine RNA-Sequenz, die der vorstehend genannten DNA-Sequenz entspricht.



Gemäß einem weiteren Aspekt der Erfindung umfaßt ein DNA-Molekül ein cDNA-Molekül, das der vorstehend genannten DNA-Sequenz entspricht.

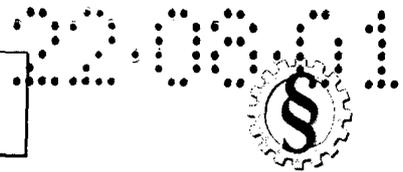
Gemäß einem weiteren Aspekt der Erfindung umfaßt ein DNA-Molekül eine
5 intronfreie DNA-Sequenz, die für ein CFTR-Mutantenpolypeptid kodiert, das die
Sequenz gemäß Fig. 1 für die Aminosäurerest-Positionen 1 bis 1480 aufweist und das
außerdem durch Nucleotidsequenzvarianten gekennzeichnet ist, die zur Deletion oder
Veränderung der Aminosäuren an mindestens einer Restposition führen, ausgewählt aus
der Gruppe bestehend aus 85, 148, 178, 455, 493, 507, 542, 549, 560, 563, 574, 1077
10 und 1092, wobei eine Veränderung an der Position 549 entweder S549R oder S549I ist.

Gemäß einem anderen Aspekt der Erfindung umfaßt ein DNA-Molekül eine
DNA-Sequenz, die für ein CFTR-Mutantenpolypeptid kodiert, das die Sequenz gemäß
Fig. 1 für die DNA-Sequenz-Positionen 1 bis 4575 aufweist und das außerdem durch
15 Nucleotidsequenzvarianten gekennzeichnet ist, die zur Deletion oder Veränderung der
DNA an mindestens einer DNA-Sequenzposition, ausgewählt aus der Gruppe bestehend
aus 129, 556 und 3659 von Fig. 1, oder, wie in Fig. 18 gezeigt, zur Deletion oder
Veränderung in der Sequenz der Introns 4, 5 und 10 an den Nucleotidpositionen, die
von 622+1, 711+1 und 1717-1 dargestellt sind, von Fig. 1 führen.

20

Gemäß einem weiteren Aspekt betrifft die Erfindung eine DNA-Sequenz, ausge-
wählt aus der Gruppe bestehend aus:

- (a) DNA-Sequenzen, die mit Abschnitten der DNA-Sequenzen von Exon/Intron-
Grenzen des genomischen CF-Gens korrespondieren;
- 25 (b) DNA-Sequenzen mit mindestens achtzehn sequentiellen Nucleotiden an Exon/
Intron-Grenzen des genomischen CF-Gens, dargestellt in Fig. 18; und
- (c) DNA-Sequenzen mit mindestens achtzehn sequentiellen Nucleotiden von Intron-
Abschnitten des genomischen CF-Gens der Fig. 18.



Gemäß einem weiteren Aspekt der Erfindung umfaßt eine gereinigte Nukleinsäuresonde eine DNA- oder RNA-Nukleotidsequenz, die den vorstehend genannten selektierten DNA-Sequenzen der Gruppen (a) bis (c) entspricht.

- 5 Gemäß einem weiteren Aspekt der Erfindung entspricht eine ein gereinigtes RNA-Molekül umfassende RNA-Sequenz der DNA-Mutantensequenz, ausgewählt aus der Gruppe der Mutantenproteinpositionen bestehend aus 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092, wobei einer Veränderung an der Position 549 entweder S549R oder S549I ist und die DNA-Mutantensequenzpositionen aus 129, 556,
10 621+1, 711+1, 1717-1 und 3659 bestehen.

Eine gereinigte Nukleinsäuresonde, die eine DNA- oder RNA-Nukleotidsequenz umfaßt, die den Mutantensequenzen der vorstehend genannten Gruppen entspricht.

- 15 Gemäß einem weiteren Aspekt der Erfindung wird ein rekombinanter Klonierungsvektor, der die DNA-Sequenzen der Mutanten-DNA und ihrer Fragmente umfaßt, ausgewählt aus der Gruppe der Mutantenprotein-Positionen bestehend aus 85, 148, 178, 455, 493, 507, 542, 549, 551, 563, 574, 1077 und 1092, wobei eine Veränderung an Position 549 entweder S549R oder S549I ist, und ausgewählt aus der Gruppe der DNA-
20 Mutantensequenzpositionen bestehend aus 129, 556, 621+1, 711+1, 1717-1 und 3659, bereitgestellt. Der Vektor wird gemäß einem Aspekt der Erfindung operativ mit einer Expressionskontrollsequenz im rekombinanten DNA-Molekül gekoppelt, so daß die selektierten DNA-Mutantensequenzen für das CFTR-Mutantenpolypeptid exprimiert werden können. Die Expressionskontrollsequenz wird aus der Gruppe selektiert be-
25 stehend aus Sequenzen, die die Expression der Gene von prokaryontischen oder eukaryontischen Zellen und ihrer Viren und deren Kombinationen kontrollieren.

Gemäß einem anderen Aspekt der Erfindung umfaßt ein Verfahren zur Produktion eines CFTR-Mutantenpolypeptids die folgenden Schritte:

- 30 (a) Kultivieren einer nicht-menschlichen Zelle, die mit dem rekombinanten Vektor für die DNA-Mutantensequenz transfiziert wurde, in einem Medium und unter



Bedingungen, die für die Expression des CFTR-Mutantenpolypeptids günstig sind, das ausgewählt ist aus der Gruppe der CFTR-Mutantenpolypeptide an den Mutantenproteinpositionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092, wobei eine Veränderung an der Position 549 entweder
5 S549R oder S549I ist, und der DNA-Mutantensequenzpositionen 129, 556, 621+1, 711+1, 1717-1 und 3659; und

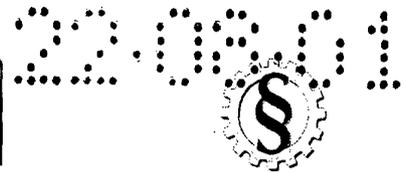
(b) Isolierung des exprimierten CFTR-Mutantenpolypeptids.

Gemäß einem weiteren Aspekt der Erfindung umfaßt ein gereinigtes Protein, das
10 aus einer menschlichen Zellmembran stammt, eine Aminosäuresequenz, die von den DNA-Mutantensequenzen kodiert wird, die ausgewählt sind aus der Gruppe der Mutantenproteinpositionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092, wobei eine Veränderung an der Position 549 entweder S549R oder S549I ist, und aus der Gruppe der DNA-Mutantensequenzpositionen 129, 556, 621+1,
15 711+1, 1717-1 und 3659, wobei das Protein, wenn es in menschlicher Zellmembran vorhanden ist, mit der Zellfunktion assoziiert wird, die die genetische Erkrankung cystische Fibrose bewirkt.

Gemäß einem weiteren Aspekt der Erfindung umfaßt ein Verfahren zum *in vitro*
20 Screening eines Subjekts, um zu bestimmen, ob es sich bei diesem Subjekt um einen CF-Träger oder einen CF-Patienten handelt, die Schritte:

Bereitstellen einer biologischen Probe des Subjekts, das einem Screening unterzogen werden soll; und Bereitstellen eines *in vitro* Assays, ausgewählt aus genetischem Assay und Immunoassay, um in der biologischen Probe die Anwesenheit mindestens
25 entweder eines CF-Mutantengens, eines CFTR-Mutantenpolypeptidprodukts oder einer Mischung davon nachzuweisen, wobei die Mutanten durch Mutationen an den DNA-Sequenzpositionen 129, 556 und 3659 der Fig. 1 oder Mutationen in der Sequenz der Introns 4, 5 und 10, wie in Fig. 18 gezeigt, an den Nukleotidpositionen, die durch 621+1, 711+1 und 1717-1 dargestellt sind, der Fig. 1 definiert sind.

30



Gemäß einem weiteren Aspekt der Erfindung umfaßt ein Kit zum Analysieren der Anwesenheit eines CF-Gens durch Immunosassaytechniken:

- (a) einen Antikörper, der spezifisch an ein Genprodukt der DNA-Mutantensequenz bindet, das ausgewählt ist aus der Gruppe der Mutantenproteinpositionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092, wobei es sich bei einer Veränderung an Position 549 entweder um S549R oder S549I handelt, und aus der Gruppe der DNA-Mutantensequenzpositionen 129, 556, 621+1, 711+1, 1717-1 und 3659;
- (b) Reagentien zum Nachweis der Bindung des Antikörpers an das Genprodukt; und wobei
- (c) die Antikörper und Reagentien jeweils in Mengen vorhanden sind, die wirksam sind, um den Immunoassay durchzuführen.

Ein weiterer Aspekt der Erfindung betrifft die Verwendung von Oligonucleotidprimern in einer Polymerase-Chainreaction (PCR), um ein ausgewähltes Exon zu amplifizieren, das eine cDNA-Sequenz der Fig. 1 aufweist, wobei diese Primer sich an Intronabschnitten in der Nähe der 5'- und 3'-Grenzen eines ausgewählten Exons von Fig. 18 befinden, ausgewählt aus der Gruppe bestehend aus Oligonucleotidprimern xi-5 und xi-3 der Tabelle 5, wobei x für die Exonnummer 1, 3, 4, 5, 6a, 6, 7 - 13, 14a, 14b, 15, 16, 17a, 17b und 18 - 24 steht.

Gemäß einem weiteren Aspekt der Erfindung umfaßt ein Tier ein heterologes Zellsystem. Das Zellsystem enthält einen rekombinanten Klonierungsvektor, der die rekombinante DNA-Sequenz enthält, die der DNA-Mutantensequenz entspricht, die die Symptome für cystische Fibrose im Lebewesen hervorrufen.

Gemäß einem weiteren Aspekt umfaßt die Erfindung in einer Polymerase-Chainreaction die ein ausgewähltes Exon einer cDNA-Sequenz der Fig. 1 amplifiziert, die Verwendung von Oligonucleotidprimern aus Intronabschnitten in der Nähe der 5'- und 3'-Grenzen des ausgewählten Exons von Fig. 18.



KURZE BESCHREIBUNG DER ZEICHNUNGEN

- 5 Fig. 1 ist die Nukleotidsequenz des CF-Gens und die Aminosäuresequenz der CFTR-Protein-Aminosäuresequenz, wobei Δ Mutationen an den Proteinpositionen 507 und 508 anzeigt.
- 10 Fig. 2 ist eine Restriktionskarte des CF-Gens und die schematische Strategie, die bezüglich des Chromosomen-Walking und -Jumping zum Gen angewendet wird.
- 15 Fig. 3 ist eine Abbildung der physischen Karte des Bereichs, der das CF-Gen einschließt und umgibt, erzeugt durch Pulsfeld-Gelelektrophorese. Die Tafeln A, B, C und zeigen Hybridisationsdaten der Restriktionsenzyme Sal I, Xho I, Sfi I und Nae I, jeweils durch repräsentative Genom- und cDNA-Sonden erzeugt, die sich über die Region erstrecken. Die für jedes Restriktionsenzym abgeleiteten physischen Karten stehen jeweils unter den Tafeln gezeigt. Eine gebräuchliche Karte für das gesamte MET-D7S8-Intervall ist in Tafel E angegeben (J.M. Rommens et al., *Am. J. Hum. Genet.* 45:932-941, 1990). Das freie Segment des Kästchens zeigt den Anteil an, der durch Chromosomen-Walking und -Jumping kloniert wurde, und der gefüllte Teil des Kästchens zeigt den Anteil an, den das CF-Transkript einnimmt.
- 20
- 25 Fig. 4A, 4B und 4C zeigen den Nachweis von konservierten Nukleotidsequenzen durch Spezieskreuzungs-Hybridisierung.
- Fig. 4D ist eine Restriktionskarte von sich überlagernden Segmenten der Sonden E4.3 und H1.6.
- 30 Fig. 5 ist eine RNA-Blot-Hybridisationsanalyse unter Verwendung von Genom- und cDNA-Sonden. Hybridisation der RNA von: A – Fibroblast mit cDNA-Sonde G-2; B – Trachea (von nicht befallenen Individuen und CF-Patienten),



Pankreas, Leber, HL60-Zell-Linie und Gehirn mit Genomsonde CF16; C –
T84-Zell-Linie mit cDNA-Sonde 10-1.

5 Fig. 6 ist der Methylierungsstatus der klonierten E4.3-Region am 5'-Ende des CF-
Gens.

Fig. 7 ist eine Restriktionskarte der CFTR-cDNA, die eine vergleichende Anordnung
der cDNA an den DNA-Genomfragmenten zeigt.

10 Fig. 8 ist eine RNA-Gelblotanalyse, die die Hybridisation eines 6,5 kb mRNA-
Transkripts durch einen Teil der CFTR-cDNA (Klon 10-1) in verschiedenen
menschlichen Geweben zeigt.

15 Fig. 9 ist eine DNA-Hybridisations-Blotanalyse, die die Hybridisation von genomi-
scher DNA durch die CFTRc-DNA-Klone zeigt, digeriert von EcoRI und
Hind III.

Fig. 10 ist ein Primer-Extensionsversuch, der die 5' und 3'-Enden der CFTR-cDNA
beschreibt.

20 Fig. 11 ist ein Hydropathiediagramm und zeigt vorhergesagte sekundäre CFTR-
Strukturen.

25 Fig. 12 ist eine Dot-Matrix-Analyse der internen Homologien im vorhergesagten
CFTR-Polypeptid.

Fig. 13 ist ein schematisches Modell des vorhergesagten CFTR-Proteins.

30 Fig. 14 ist ein schematisches Diagramm von Restriktionsfragmentlängen-Polymorphis-
men (RFLPs), die eng mit dem CF-Gen gekoppelt sind, wobei das umgekehrte
Dreieck die Lage der Deletion der 3 Basenpaare von F508 anzeigt.

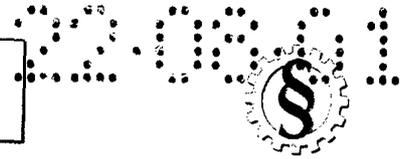


Fig. 15 stellt eine vergleichende Anordnung der meisten konservierten Segmente der extendierten NBFs der CFTR mit vergleichbaren Bereichen anderer Proteine dar.

5

Fig. 16 ist die DNA-Sequenz um die F508-Deletion.

Fig. 17 ist eine Darstellung des Nucleotidsequenzierungsgels, die die DNA-Sequenz an der F508-Deletion zeigt.

10

Fig. 18 ist die Nukleotidsequenz von Intronabschnitten und kompletten Exons des genomischen CF-Gens für 27 Exons, die identifiziert und hintereinander mit 1 bis 24 numeriert wurden, mit zusätzlichen Exons 6a, 6b, 14a, 14b und 17a, 17b der cDNA-Sequenz von Fig. 1.

15

Fig. 19 zeigt die Ergebnisse der Amplifikation der genomischen DNA unter Verwendung des die Intron-Oligonucleotide einschließenden Exons 10;

Fig. 20 zeigt die Separation der amplifizierten genomischen DNA-Produkte einer CF-Familie mittels Gelelektrophorese; und

20

Fig. 21 ist eine Restriktionskartierung von klonierten Intron- und Exonabschnitten der genomischen DNA wobei die Introns und Exons in Fig. 18 identifiziert sind.

25



DETAILLIERTE BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORMEN

1. DEFINITIONEN

5

Um den Überblick über die verschiedenen Ausführungsformen der Erfindung und das Verständnis verschiedener Elemente und Bestandteile, die bei der Erfindung und ihrer Durchführung verwendet wurden, zu erleichtern, werden die in der Beschreibung der Erfindung verwendeten Ausdrücke wie folgt definiert:

10 CF – cystische Fibrose

CF-Träger – eine anscheinend gesunde Person, deren Chromosomen ein CF-Mutantengen enthalten, das auf die Nachkommen dieser Person übertragen werden kann.

15 CF-Patient – eine Person, die auf jedem Chromosom ein CF-Mutantengen trägt, und die somit die klinischen Symptome der zystischen Fibrose zeigt.

CF-Gen – das Gen, dessen Mutantenformen mit der Krankheit cystische Fibrose zusammenhängen. Diese Definition soll die verschiedenen existierenden Sequenzpolymorphismen einschließen, wobei Nucleotid-Substitutionen in der Gensequenz die grundlegende Funktion des Genprodukts nicht beeinflussen.

20 Dieser Ausdruck bezieht sich in erster Linie auf eine isolierte Kodierungssequenz, kann aber auch einige oder alle der flankierenden Regulations-elemente und/oder Introns einschließen.

Genomisches CF-Gen – das CF-Gen, das flankierende Regulationselemente und/oder Introns an den Exongrenzen des CF-Gens einschließt.

25 CF-PI – an cystischer Fibrose mit Bauchspeicheldrüsen-Unterfunktion leidende Personen, die größte klinische Untergruppe der cystische Fibrose-Patienten, gekennzeichnet durch eine Unterfunktion der Bauchspeicheldrüsen-Funktion.

30 CF-PS – an cystischer Fibrose leidende Personen mit normaler Bauchspeicheldrüsen-Funktion, eine klinische Untergruppe von cystische Fibrose-Patienten mit normaler Bauchspeicheldrüsen-Funktion, die zur normalen Verdauung der Nahrung fähig sind.



- CFTR – cystische Fibrose-Leitfähigkeitsregulator-Transmembranprotein, kodiert durch das CF-Gen. Diese Definition schließt Protein, wie von menschlichen oder Tier-Quellen isoliert, wie von rekombinanten Organismen erzeugt und wie chemisch oder enzymatisch synthetisiert ein. Diese Definition soll die verschiedenen polymorphen Formen des Proteins einschließen, wobei Aminosäure-Substitutionen in den verschiedenen Regionen der Sequenz die grundlegende Funktion des Proteins oder sein Hydropathiediagramm oder seine sekundäre oder tertiäre Struktur nicht beeinflussen.
- 5
- 10 DNA – Zur Identifikation der Basen wird die Standardnomenklatur verwendet.
Intronfreie DNA – ein DNA-Stück, das keine nicht-kodierenden internen Segmente aufweist, beispielsweise cDNA.
IRP-Locussequenz – (Protooncogen int-1-betreffend), ein Gen, das sich in der Nähe des CF-Gens befindet.
- 15 CFTR-Mutant – ein Protein, das weitgehend analog mit dem CFTR ist, was seine primäre, sekundäre und tertiäre Struktur betrifft, aber in dem eine kleine Zahl von Aminosäuresubstitutionen und/oder -deletionen und/oder -insertionen zu einer Störung seiner Grundfunktion führen, so daß Organismen, deren Epithelzellen CFTR-Mutanten anstelle von CFTR exprimieren, die Symptome der zystischen Fibrose zeigen.
- 20 mCF – ein Mäusegen, das zum menschlichen CF-Gen ortholog ist.
NBFs – Nukleotid (ATP) -Bindungsfalten
ORF – Offener Leseraster
PCR – Polymerase-Chainreaction
- 25 Protein – es wird die Standard-Einbuchstabenomenklatur verwendet, um die Aminosäuren zu identifizieren.
R-Domäne – eine hoch aufgeladene Zytoplasmadomäne des CFTR-Proteins
RSV – Rous Sarcoma Virus
SAP – oberflächenaktives Protein
- 30 RFLP – Restriktionsfragmentlängen-Polymorphismus



507 CF-Mutantengen – das CF-Gen, das eine DNA-Basenpaarmutation an der 506 oder
507 Proteinposition der cDNA des CF-Gens enthält.

507-DNA-Mutantensequenz – äquivalenter Ausdruck für das 507 CF-Mutantengen.

507 CFTR-Mutantenprotein oder CFTR-Mutantenprotein-Aminosäuresequenz oder
5 CFTR-Mutantenpolypeptid – das CFTR-Mutantenpolypeptid, in dem eine
Aminosäuren-Deletion an der 506 oder 507 Isoleucinprotein-Position des
CFTR auftritt.

2. ISOLIERUNG DES CF-GENS

10

Mittels des Chromosomen-Walking, -Jumping und der cDNA-Hybridisation
wurden DNA-Sequenzen, die > 500 Kilobasenpaare (kb) umfassen, aus einer Region
auf dem langen Arm des menschlichen Chromosoms 7 isoliert, der das cystische Fibrose
(CF) -Gen enthält. Dieses Verfahren wird in den vorstehend genannten mit-anhängigen
15 US-Patentanmeldungen detailliert offenbart. Um das Verstehen der Isolierung des CF-
Gens und der Identifikation anderer Mutationen, beispielsweise an den 85, 148, 1178,
455, 493, 507, 542, 549, 560, 563, 574, 1077 und 1092 Aminosäurerest-Positionen, zu
erleichtern, wird die Technik hierin noch einmal beschrieben. Es wurden in dieser
Region mehrere transkribierte Sequenzen und konservierte Segmente identifiziert. Eines
20 davon entspricht dem CF-Gen und umfaßt ungefähr 250 kb der Genom-DNA. Sich
überlagernde komplementäre DNA (cDNA)-Klone wurde aus Epithelzellbibliotheken
mit einem genomischen DNA-Segment isoliert, das einen Teil des cystischen Fibrose-
gens enthält. Die Nucleotidsequenz der isolierten cDNA ist in den Fig. 1 bis 18 gezeigt.
In jeder Reihe der jeweiligen Sequenzen stellt die untere Reihe eine Standardnomen-
25 klatur-Liste der Nucleotidsequenz dar. Die obere Reihe in den jeweiligen Sequenzreihen
ist die Standard-Einbuchstabenomenklatur der Aminosäure, die dem jeweiligen Codon
entspricht.

Demgemäß ergab die Isolierung des CF-Gens ein cDNA-Molekül, das eine DNA-
30 Sequenz umfaßt, die ausgewählt ist aus der Gruppe bestehend aus:



- (a) DNA-Sequenzen, die der DNA-Sequenz von Fig. 1 entsprechen, von Aminosäurerest-Position 1 bis Position 1480;
- (b) DNA-Sequenzen, die für normales CFTR-Polypeptid, das die Sequenz gemäß Fig. 1 aufweist, kodieren, für die Aminosäurerest-Positionen von 1 bis 1480;
- 5 (c) DNA-Sequenzen, die einem Fragment der Sequenz von Fig. 1 entsprechen und mindestens 16 sequentielle Nucleotide zwischen den Aminosäurerest-Positionen 1 und 1480 einschließen;
- (d) DNA-Sequenzen, die mindestens 16 Nucleotide umfassen und für ein Fragment der Aminosäuresequenz von Fig. 1 kodieren; und
- 10 (e) DNA-Sequenzen, die für ein Epitop kodieren, kodiert von mindestens 18 sequentiellen Nucleotiden in der Sequenz von Fig. 1 zwischen den Aminosäurerest-Positionen 1 und 1480.

Gemäß der Erfindung liefert die Isolierung anderer Mutationen im CF-Gen auch
15 ein cDNA-Molekül, das eine DNA-Sequenz umfaßt, die ausgewählt ist aus der Gruppe bestehend aus:

- a) DNA-Sequenzen, die der DNA-Sequenz, die für CFTR-Mutantenpolypeptide kodiert, die durch mit zystischer Fibrose assoziierter Aktivität in menschlichen Epithelzellen gekennzeichnet sind, oder der DNA-Sequenz von Fig. 1 für die
20 Aminosäurerest-Positionen 1 bis 1480, aber weiterhin gekennzeichnet durch eine Basenpaar-Mutation, die zur Deletion oder Veränderung einer Aminosäure an den Restpositionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092 führt, entsprechen;
- b) DNA-Sequenzen, die Fragmenten des Mutantenabschnitts der Sequenz des Abschnitts a) entsprechen und die mindestens sechzehn Nucleotide enthalten;
- 25 c) DNA-Sequenzen, die mindestens sechzehn Nucleotide umfassen und für ein Fragment der Aminosäuresequenz kodieren, kodiert vom Mutantenabschnitt der DNA-Sequenz des Abschnitts a); und
- d) DNA-Sequenzen, die für ein Epitop kodieren, kodiert von mindestens 18
30 sequentiellen Nucleotiden im Mutantenabschnitt der Sequenz der DNA von Abschnitt a).



Transkripte in der Größenordnung von ungefähr 6500 Nucleotiden sind in b-
efallenen Geweben von CF-Patienten nachweisbar. Aufgrund der isolierten Nucleotid-
sequenz besteht das vorhergesagte Protein aus zwei ähnlichen Regionen, von denen jede
5 eine erste Domäne, deren Eigenschaften auf einen Membran-Zusammenhang schließen
lassen, und eine zweite Domäne, von der man annimmt, daß sie an der ATP-Bindung
beteiligt ist, enthält.

Eine 3 bp-Deletion, die zum Fehlen eines Phenylalaninrests im Zentrum der ersten
10 vorhergesagten Nucleotid-Bindungsdomäne (Aminosäureposition 508 des CF-Gen-
produkts) führt, wurde in CF-Patienten nachgewiesen. Diese Mutation der normalen
DNA-Sequenz von Fig. 1 entspricht ungefähr 70 % der Mutationen bei cystische
Fibrose-Patienten. Extendierte Haplotypdaten aufgrund von DNA-Markern, die eng mit
dem mutmaßlichen Krankheitsgen gekoppelt sind, legen nahe, daß der Rest des CF-
15 Mutantengenpools aus multiplen, unterschiedlichen Mutationen besteht. Dies wird nun
durch die Erfindung, beispielsweise an der 506 oder 507 Proteinposition, erklärt. Eine
kleine Gruppe dieser letztgenannten Mutantenallele (ungefähr 8 %) können zu einer
andauernden normalen Bauchspeicheldrüsen-Funktion bei einer Untergruppe der
Patienten, deren Bauchspeicheldrüse normal funktioniert, beitragen.

20

2.1 CHROMOSOMEN-WALKING ODER -JUMPING

Große Mengen der DNA, die die D7S122- und D7S340-Kopplungsbereiche von
Rommens et al., s.o., umgeben, wurden nach Kandidatengensequenzen durchsucht.
25 Zusätzlich zu herkömmlichen Chromosomen-Walking-Verfahren wurden Chromo-
somen-Jumping-Verfahren angewandt, um die Suche zu beschleunigen. Von jedem
Jumping-Endpunkt konnte ein neues bidirektionales Walking initiiert werden. Aufein-
anderfolgende Walkings, die von „nicht klonierbaren“ Regionen aufgehalten wurden,
die im Säuger-Genom häufig anzutreffen sind, konnten durch das Chromosomen-
30 Jumping umgangen werden.



Die verwendete Chromosomen-Jumping-Bibliothek wurde bereits beschrieben [Collins et al., *Science* 235, 1046 (1987); Ianuzzi et al., *Am. J. Hum. Genet.* 44, 695 (1989)]. Die ursprüngliche Bibliothek wurde aus einem Präparations-Pulsfeldgel hergestellt und sollte teilweise EcoR1-Fragmente von 70 – 130 kb enthalten; die
5 folgenden Erfahrungen mit dieser Bibliothek zeigten, daß auch kleinere Fragmente anwesend waren, und es wurden Jumping-Weiten von 25 – 110 kb gefunden. Die Bibliothek wurde auf sup⁻ Wirts-MC1061 aufgestrichen und Standard-Screening-verfahren unterzogen [Maniatis et al.]. Positive Klone wurden zu pBRΔ23Ava sub-
10 kloniert und der Anfang und das Ende des Jumping wurden durch EcoR1- und Ava I-Digestion identifiziert, wie in Collins, *Genome Analysis: A practical approach* (IRL; London, 1988), S. 73-94, beschrieben wird. Für jeden Klon wurde ein Fragment am Ende des Jumping untersucht, um dessen Lage auf dem Chromosom 7 zu bestätigen. Die benachbarte Chromosomenregion, die vom Chromosomen-Walking und -Jumping abgedeckt wurde, betrug etwa 250 kb. Die Jumping-Richtung wurde durch sorgfältige
15 Wahl der Sonden beeinflusst, wie von Collins et al. und Ianuzzi et al., s.o., beschrieben wurde. Die gesamte klonierte Region, einschließlich der Sequenzen, die mittels der cDNA des CF-Gens isoliert wurden, umfaßt ungefähr 500 kb.

Die schematische Darstellung der Strategie des Chromosomen-Walking und
20 -Jumping ist in Fig. 2 dargestellt. CF-Gen-Exons sind in dieser Figur mit römischen Ziffern bezeichnet. Horizontale Linien über der Karte zeigen die Walking-Schritte an, wogegen die Bögen über der Karte Jumping-Schritte anzeigen. Die Figur schreitet in jeder der sechs Stufen von links nach rechts voran, wobei ihre Enden auf 7cen und 7ter gerichtet sind, wie angezeigt. Die Restriktionskarte für die Enzyme EcoRI, HindIII und
25 BamHI ist über der durchgezogenen Linie gezeigt, und umfaßt die gesamte klonierte Region. Restriktionsstellen, die mit Pfeilen anstelle von vertikalen Linien bezeichnet sind, zeigen Stellen an, deren Position nicht eindeutig ist. Zusätzliche Restriktionsstellen für andere Enzyme sind unter der Linie gezeigt. Lücken in der klonierten Region sind durch || angezeigt. Diese treten nur in dem Abschnitt auf, der von cDNA-Klonen
30 auf dem CF-Transkript nachgewiesen wurde. Diese Spalten sind aufgrund der Pulsfeld-Kartierung der Region wahrscheinlich nicht groß. Die Walking-Klone, wie von



horizontalen Pfeilen über der Karte angezeigt, weisen die Richtung des Pfeils auf, der den Walkingfortschritt anzeigt, der mit jedem Klon erhalten wurde. Cosmidklone beginnen mit dem Buchstaben c; alle anderen Klone sind Phagen. Cosmid CF26 erwies sich als Chimäre; der schraffierte Bereich ist von einem anderen Genomfragment auf einem anderen Chromosom abgeleitet. Die römischen Zahlen I bis XXIV zeigen die Lage des Exons auf dem CF-Gen an. Die über der Line gezeigten horizontalen Kästchen sind Sonden, die während der Versuche verwendet wurden. Drei dieser Sonden stellen das unabhängige Subklonieren von Fragmenten dar, die zuvor identifiziert worden waren, um Polymorphismen in dieser Region nachzuweisen: H2.3A entspricht der Sonde XV2C (X. Estivill et al., *Nature*, 326: 840 (1987)), Sonde E1 entspricht KM19 (Estivill, s.o.) und Sonde E4.1 entspricht Mp6d.9 (X. Estivill et al., *Am. J. Hum. Genet.* 44, 704 (1989)). G-2 ist ein Subfragment von E6, das eine transkribierte Sequenz nachweist. R161, R159 und R160 sind thetische Oligonucleotide, die aus Teilen der IRP-Locussequenz konstruiert wurden [B.J. Wainwright et al., *EMBO J.* 7: 1743 (1988)], die die Lage dieses Transkripts auf der Genomkarte anzeigen.

Da die beiden unabhängig voneinander isolierten DNA-Marker D7S122 (pH131) und D7S340 (TM58) nur ungefähr 10 kb voneinander entfernt waren (Fig. 2), gingen die Walkings und Jumpings praktisch von einem einzigen Punkt aus. Dann wurde die Richtung des Walking und Jumping bezüglich MET und D7S8 mit dem Kreuzen von mehreren selten schneidenden Endonucleaseerkennungs-Restriktionsstellen (wie denen für Xho I, Nru I und Not I, siehe Fig. 2) und mit Bezug auf die physische Karte langer Fragmente von J. M. Rommens et al., *Am. J. Hum. Genet.*, in Druck; A. M. Poustka et al., *Genomics* 2, 337 (1988); M.L. Drumm et al., *Genomics* 2, 346 (1988) bestimmt. Die Pulsfeldkartierungs-Daten zeigten auch, daß die Not I-Stelle, die von den Erfindern der vorliegenden Erfindung identifiziert wurde (siehe Fig. 2, Position 113 kb) mit der zuvor gefundenen übereinstimmte, die mit dem IRP-Locus in Zusammenhang gebracht wurde (Estivill et al., 1987, s.o.). Da anschließende genetische Studien zeigten, daß CF sehr wahrscheinlich zwischen IRP und D7S8 liegt [M. Farrall et al., *Am. J. Hum. Genet.* 43, 471 (1988), B.S. Kerem et al. *Am. J. Hum. Genet.* 44, 827 (1989)], konzentrierte man sich bei den weiteren Walking- und Jumpingversuchen ausschließlich auf das Klonieren



dieses Intervalls. Es ist jedoch von Vorteil, daß andere Kodierungsregionen, wie in Fig. 2 identifiziert, beispielsweise G-2, CF14 und CF16, lokalisiert und gründlich untersucht worden waren. Diese umfassenden Untersuchungen dieser anderen Regionen zeigten aufgrund von genetischen Daten und Sequenzanalysen, daß es sich dabei nicht um das CF-Gen handelte. Angesichts des fehlenden Wissens um die Lage des CF-Gens und seine Eigenschaften, brachte die umfassende und zeitaufwendige Überprüfung der in der Nähe liegenden angenommenen Kodierungsregionen keinen Fortschritt bei der Suche nach dem CF-Gen. Diese Untersuchungen waren jedoch nötig, um die Möglichkeit auszuschließen, daß das CF-Gen sich in diesen Regionen befindet.

Es wurde gefunden, daß drei Regionen im 280 kb-Segment in den amplifizierten Genombibliotheken, die zu Anfang verwendet wurden, nicht leicht zu erhalten waren. Diese weniger klonierbaren Regionen wurden in der Nähe der DNA-Segmente H2.3A und X.6 und unmittelbar hinter Cosmid cW44 an den Positionen 75 – 100 kb, 205 – 225 kb bzw. 275 – 285 kb in Fig. 2 lokalisiert. Es wurde gefunden, daß die rekombinanten Klone in der Nähe von H2.3A sehr instabil waren, wobei nach nur wenigen Bakterienkultur-Durchgängen dramatische Umgruppierungen stattfanden. Um die resultierenden Lücken zu füllen, wurden primäre Walking-Bibliotheken unter Verwendung spezieller Wirts-Vektorsysteme konstruiert, von denen berichtet worden war, daß sie die Vermehrung von instabilen Sequenzen ermöglichen [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)]. Obwohl die Region in der Nähe von Cosmid cW44 immer noch nicht erforscht wurde, gelang es, die Region in der Nähe von X.6 mit diesen Bibliotheken zu retten.

2.2 KONSTRUKTION DER GENOMBIBLIOTHEKEN

Die Genombibliotheken wurden gemäß Vorgehensweisen konstruiert, die in Maniatis et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 1982) beschrieben sind und sind in



Tabelle I aufgeführt. Dies schließt acht Phagenbibliotheken ein, von denen eine von T. Maniatis [Fritsch et al., *Cell*, 19:959 (1980)] bereitgestellt wurde: der Rest wurde als Teil dieser Arbeit gemäß dem in Maniatis et al., s.o., beschriebenen Verfahren konstruiert. Vier Phagenbibliotheken wurden in λ DASH (im Handel von Stratagene erhältlich) und drei in λ FIX (im Handel von Stratagene erhältlich) kloniert, mit Vektorarmen, die vom Hersteller bereitgestellt wurden. Eine λ DASH-Bibliothek wurde aus Sau3A-teildigierter DNA aus einem Mensch/Hamster-Hybrid konstruiert, das das menschliche Chromosom 7 (4AF/102/K015) enthielt [Rommens et al., *Am. J. Hum. Genet* 43, 4(1988)] und andere Bibliotheken wurden aus Sau3A-teil-, BamHI-ganz- oder EcoRI-ganzdigierter humaner peripherer Blut- oder lymphoblastoider DNA konstruiert. Um einen Verlust von instabilen Sequenzen zu verhindern, wurden fünf der Phagen-Bibliotheken auf dem Rekombinations-Mangel Wirt DB1316 (*recD*⁻), CES 200 (*recBC*⁻) [Wyman et al., s.o., Wertman et al., s.o., Wyman et al., s.o.] oder TAPÜ90 [Patterson et al., *Nucleic Acids. Res.* 15:6298 (1987)] vermehrt. Dann wurden drei Cosmid-Bibliotheken konstruiert. In einer wurde der Vektor pCV108 [Lau et al., *Proc. Natl. Acad. Sci. USA* 80: 5225 (1983)] verwendet, um teildigierete (Sau 3A-) DNA aus 4AF/102/K015 zu klonieren [Rommens et al., *Am. J. Hum. Genet.* 43:4 (1988)]. Eine zweite Cosmid-Bibliothek wurde durch Klonieren von teildigierter (Mbo I) menschlicher, lymphoblastoider DNA in den Vektor pWE-IL2R hergestellt, hergestellt durch Insertion von Promotor-getriebener RSV (Rous Sarcoma Virus) -cDNA für die α -Kette des Interleukin-2-Rezeptors (erhalten von M. Fordis und B. Howard) anstelle des neo-Resistenzgens von pWE15 [Wahl et al., *Proc. Natl. Acad. Sci. USA* 84:2160 (1987)]. Zusätzlich wurde eine Mbo I-Cosmid-Teilbibliothek im Vektor pWE-IL2-Sal hergestellt, hergestellt durch Insertion eines Sal I-Linkers in die Bam HI-Klonierungsstelle von pWE-EL2R (M. Drumm, nicht-veröffentlichte Daten); dies ermöglicht die Anwendung der Teilfüllungs-Technik, um Sal I- und Mbo I-Enden zu ligieren, was Tandem-Insertionen verhindert [Zabarovsky et al., *Gene* 42:19 (1986)]. Cosmid-Bibliotheken wurden in *E. coli* Wirtstämmen DH1 oder 490A vermehrt [M. Steinmetz, A. Winoto, K. Minard, L. Hood, *Cell* 28, 489 81982)].

30



TABELLE 1
GENOMBIBLIOTHEKEN

Vektor	Human-DNA-Quelle	Wirt	Komplexität	Ref.
λ Charon 4A	HaeII/AluI-teildigerierte vollständige, menschliche Leber-DNA	LE392	3×10^6 (amplifiziert)	Lawn et al., 1980
pCV108	Sau3a-teildigerierte DNA von 4AF/KO15	DK1	3×20^6 (amplifiziert)	
λ dash	Sau3A-teildigerierte DNA von 4AF/KO15	LE392	1×10^6 (amplifiziert)	
λ dash	Sau3A-teildigerierte vollständige menschliche periphere Blut-DNA	DB1316	$1,5 \times 10^6$	
λ dash	BamHI-digerierte vollständige menschliche periphere Blut-DNA	DB1316	$1,5 \times 10^6$	
λ dash	EcoRI-teildigerierte vollständige, menschliche periphere Blut-DNA	DB1316	8×10^6	
λ FIX	MboI-teildigerierte, menschliche lymphoblastoide DNA	LE392	$1,5 \times 10^6$	
λ FIX	MboI-teildigerierte, menschliche lymphoblastoide DNA	CE200	$1,2 \times 120^6$	
λ FIX	MboI-teildigerierte, menschliche lymphoblastoide DNA	TAP90	$1,3 \times 10^6$	
pWE-IL2R	MboI-teildigerierte, menschliche Lymphoblasten-DNA	490A	5×10^5	
pWE-IL2R-Sal	MboI-teildigerierte, menschliche lymphoblastoide DNA	490A	$1,2 \times 10^6$	
λ Ch3A (jumping)	Δ lac EcoRI-teildigerierte, menschliche lymphoblastoide DNA (24 – 1210 kb)	MC1061	3×10^6	Collins et al. (s.o.) und Ianuzzi et al. (s.o.)



Drei dieser Phagen-Bibliotheken wurden in *E. coli*-Bakterienstamm LE392 vermehrt und amplifiziert. Vier darauffolgende Bibliotheken wurden auf die Rekombinations-Mangel-Wirte DB1316 (*recD*⁻) oder CES200 (*rec BC*⁻) [Wyman 1985, s.o.; Wertman 1986, s.o.; und Wyman 186. s.o.] oder in einem Fall auf TAP90 [T.A. Patterson und M. Dean, *Nucleic Acids Research* 15, 6298 (1987)] aufgestrichen.

Einzelkopie-DNA-Segmente (ohne Wiederholungselemente) in der Nähe der Enden jedes Phagen- oder Cosmid-Inserts wurden gereinigt und als Sonden für das Bibliotheken-Screening verwendet, um sich überlagernde DNA-Fragmente durch Standardverfahren zu isolieren (Maniatis et al., s.o.).

$1 - 2 \times 10^6$ Phagenklone wurden auf 25 - 30 150 mm-Petrischalen mit dem geeigneten bakteriellen Indikatorwirt gestrichen und bei 37° 10 - 16 h lang inkubiert. Duplikat-, „Lifts“ wurden von jeder Platte mit Nitrocellulose- oder Nylon-Membranen hergestellt, und unter Bedingungen, die bereits beschrieben wurden, prehybridisiert und hybridisiert [Rommens et al., 1988, s.o.]. Sonden wurden mit ³²P bezüglich einer spezifischen Aktivität von $> 5 \times 10^8$ cpm/μg mittels des Zufalls-Primingverfahrens markiert [A.P. Feinberg und B. Vogelstein, *Anal. Biochem.* 132, 6 (1983)]. Die Cosmid-Bibliothek wurde auf Ampicillin-haltigen Platten verstrichen und auf ähnliche Weise einem Screening unterzogen.

DNA-Sonden, die starke Hintergrundsignale aussandten, konnten häufig mit größerem Erfolg verwendet werden indem man die abgekochte Sonde mit 250 μg/ml zerkleinerter, denaturierter Placenta-DNA 60 Minuten lang vorerhitzte, bevor man die Sonde in den Hybridisationsbeutel gab.

Für jeden Walking-Schritt wurde die Identität des klonierten DNA-Fragments durch Hybridisation mit einer Somazellhybrid-Tafel, um seine chromosomale Lage zu bestimmen, und durch Restriktionskartierung und Southern Blot-Analyse, um seine Co-linearität mit dem Genom zu bestätigen, bestimmt.



Die gesamte kombinierte klonierte Region der isolierten genomischen DNA-Sequenzen und der sich überlagernden cDNA-Klone umfasste > 500 kb. Um zu gewährleisten, daß die durch die Chromosomen-Walking- und -Jumping-Verfahren isolierten DNA-Segmente mit der Genomsequenz colinear waren, wurde jedes Segment untersucht durch:

- (a) Hybridisationsanalyse mit somatischen Mensch/Nager-Hybrid-Somazelllinien, um den Standort auf dem Chromosom 7 zu bestätigen,
- (b) Pulsfeld-Gelelektrophorese und
- (c) Vergleich der Restriktionskarte der klonierten DNA mit der der genomischen DNA.

Demgemäß wurden aus jedem rekombinanten Phagen- und Cosmidklon Einzelkopien von menschlichen DNA-Sequenzen isoliert und bei jeder dieser Hybridisationsanalysen, die nach dem Verfahren von Maniatis et al, s.o., durchgeführt wurden, als Sonden verwendet.

Obwohl die Mehrzahl der Phagen- und Cosmidisolate korrekte Walking- und Jumping-Klone darstellten, stammten einige aus der Klonierung von Artefakten oder aus Kreuzungshybridisations-Sequenzen aus anderen Regionen im menschlichen Genom oder, in Fällen, wo die Bibliotheken aus einer Mensch/Hamster-Hybridzelllinie stammt, aus dem Hamstergenom. Die Bestätigen der korrekten Lokalisierung war besonders wichtig bei Klonen, die durch Chromosomen-Jumping isoliert worden waren. Viele Jumping-Klone wurden betrachtet und lieferten nicht-schlüssige Informationen, was von der Untersuchung des Gens ablenkte.

2.3 BESTÄTIGUNG DER RESTRIKTIONSKARTE

Eine weitere Bestätigung der gesamten physischen Karte der sich überlagernden Klone wurde durch eine Restriktionskartierungsanalyse langer Fragmente mittels der Pulsfeld-Gelelektrophorese erhalten (J.M. Rommens et al., *A. J. Hum. Genet.*, in Druck, A. M. Poustka et al., 1988, s.o., M.L. Drumm et al., 1988, s.o.).



Figuren 3A bis 3E veranschaulichen die Ergebnisse der Restriktionskartierungs-
analyse von langen Fragmenten, wobei eine schematische Darstellung der Region in
Tafel E angegeben ist. Die DNA aus der Mensch/Hamster-Zelllinie 4AF/102/K015
5 wurde mit den Enzymen (A) Sa1 I, (B) Xho I, (C) Sfi I und (D) Nae I digeriert, mittels
Pulsfeld-Gelelektrophorese separiert und auf Zetaprobe™ (BioRad) übertragen. Für
jedes Enzym wurde ein einzelner Blot nacheinander mit den Sonden hybridisiert, die
unter jeder der Tafeln der Fig. A bis D angegeben sind, indem man den Blot zwischen
den Hybridisationen abstreifte. Die Symbole für jedes Enzym der Fig. 3E sind: A, Nae
10 I; B, Bss HII; F, Sfi I; L, Sa1 I; M, Mlu I; N, Not I; R, Nru I; und X, Xho I. C entspricht
der Kompressionszonenregion des Gels. Die DNA-Präparationen, die Restriktions-
digestion und die Kreuzungs-Feld-Gelelektrophoreseverfahren wurden beschrieben
(Rommens et al., in Druck, s.o.) Die Gele in Fig. 3 wurden in 0,5X TBE für (A), (B)
und (C) bei 7 Volt/cm 20 Stunden lang durchgeführt, wobei linear um 10 – 40 Sekunden
15 gesteigert umgeschaltet wurde, und für (D) bei 8 Volt/cm 20 Stunden lang durchgeführt,
wobei linear um 50 – 150 Sekunden gesteigert wurde. Schematische Interpretationen der
Hybridisationsmuster sind unter jeder Tafel angegeben. Die Fragmentlängen sind in
Kilobasen angegeben und ihre Größe wurde durch Vergleich mit oligomerisierten
Bakteriophagen -λDNA und *Saccharomyces cerevisiae*-Chromosomen bestimmt.

20 H4.0, J44, EG1.4 sind Genomsonden, die aus den Walking- und Jumping-
Versuchen erhalten wurden (siehe Fig. 2). J30 wurde durch vier aufeinanderfolgende
Sprünge aus D7S8 isoliert (Collins et al., 1987, s.o.; Ianuzzi et al., 1989, s.o.; M. Dean
et al., zur Veröffentlichung eingereicht). 10-1, B.75 und CE1.5/1.0 sind cDNA-Sonden,
25 die verschiedene Regionen des CF-Transkripts abdecken: 10-1 enthält Exons I – VI,
B.75 enthält Exons V – XII und CE1.5/1.0 enthält Exons XII – XXIV. In Fig. 3E ist
eine zusammengesetzte Karte des gesamten MET – D7S8-Intervalls gezeigt. Die leere
Fläche zeigt das Segment an, das durch Walking und Jumping kloniert wurde, und der
geschlossene Teil zeigt die Region an, die vom CF-Transkript abgedeckt wird. Die
30 CpG-reiche Region, die mit dem D7S23-Locus in Zusammenhang gebracht wird (Esti-
vill et al., 1987, s.o.) liegt an der Not I-Stelle, die in Klammern angegeben ist. Diese und



andere Stellen, die in Klammern oder eckigen Klammern angegeben sind, sind keine Einschnitte in 4AF/102/K015, sondern wurden in menschlichen Lymphoblast-Zelllinien entdeckt.

5 2.4 IDENTIFIKATION DES CF-GENS

Aufgrund der Ergebnisse der Restriktionskartierung von langen Fragmenten, die vorstehend detailliert erläutert wurde, wurde nachgewiesen, daß das gesamte CF-Gen in einem 380 kb Sal I-Fragment enthalten ist. Die vergleichende Anordnung der Restriktionsstellen, die aus der Pulsfeld-Gelanalyse abgeleitet wurden, mit denen, die in den sich teilweise überlagernden genomischen DNA-Klonen identifiziert wurden, zeigte, daß die Größe des CF-Gens ungefähr 250 kb beträgt.

Das informativste Restriktionsenzym, das dazu diente, die Karte des klonierten DNA-Fragments und die Restriktionskarte der langen Fragmente vergleichend anzuordnen, war Xho I; alle 9 Xho I-Stellen, die mit den rekombinanten DNA-Klonen identifiziert wurden, schienen für mindestens eine teilweises Spaltung der genomischen DNA anfällig zu sein (vergleiche Karten in Fig. 2 und 3B). Außerdem identifizierte die Hybridisationsanalyse mit Sonden, die von dem 3'-Ende des CF-Gens abgeleitet worden waren, 2 SfiI-Stellen und bestätigte die Position auf einer vorhergesagten Nae I-Stelle.

Diese Erkenntnisse stützten weiterhin die Schlußfolgerung, daß die DNA-Segmente, die durch die Chromosomen-Walking- und -Jumping-Verfahren isoliert worden waren, mit den genuinen Sequenzen colinear waren.

25 2.5 KRITERIEN FÜR DIE IDENTIFIKATION

Ein positives Ergebnis auf der Grundlage eines oder mehrerer der folgenden Kriterien ließ darauf schließen, daß ein kloniertes DNA-Segment Kandidatengensequenzen enthalten kann:



- (a) Nachweis von Kreuzungshybridisations-Sequenzen in anderen Spezies (da viele Gene eine evolutionäre Konservierung zeigen),
- (b) Identifikation von CpG-Inseln, die häufig das 5'-Ende von Wirbeltiergenen markieren [A.P. Bird, *Nature*, 321, 209 (1986); M. Gardiner-Garden und M. Frommer, *J. Mol. Biol.* 196, 261 (1987)],
- (c) Untersuchung von möglichen mRNA-Transkripten in befallenen Geweben von CF-Patienten,
- (d) Isolierung von entsprechenden cDNA-Sequenzen,
- (e) Identifikation von offenen Leserahmen durch direktes Sequenzieren von klonierten DNA-Segmenten.

Die Spezieskreuzungs-Hybridisation zeigte eine starke Sequenzkonservierung zwischen menschlicher und Rinder-DNA, wenn CF14, E4.3 und H1.6 als Sonden verwendet wurden, die Ergebnisse sind in den Fig. 4A, 4B und 4C gezeigt.

Genomische Menschen-, Rinder-, Mäuse-, Hamster- und Hühner-DNAs wurden mit Eco RI (R), Hind III (H) und Pst I (P) digeriert, elektrophoretisiert und auf Zetabind[®] (BioRad) geblottet. Die Hybridisationsverfahren von Rommens et al., 1988, s.o., wurden mit der wirkungsvollsten Wäsche bei 55°C, 0,2X SSC und 0,1 % SDAS angewendet. Die Sonden, die für die Hybridisation verwendet wurden, in Fig. 4, enthielten: (A) vollständiges Cosmid CF14, (B) E4.3, (C) H1.6. Im Schema von Fig. (D) zeigt die schattierte Region die Fläche der Spezieskreuzungs-Konservierung an.

Die Tatsache, daß verschiedene Subgruppen von Banden in Rinder-DNA mit diesen beiden sich überlagernden DNA-Segmenten (H1.6 und E4.3) entdeckt wurden, ließ darauf schließen, daß die konservierten Sequenzen sich an den Grenzen der überlagerten Region befanden (Fig. 4(D)). Wenn diese DNA-Segmente verwendet wurden, um RNA-Transkripte aus einer Reihe von Geweben nachzuweisen, wurde kein Hybridisationssignal nachgewiesen. Um zu versuchen, die Kreuzungshybridisations-Region zu verstehen, und um mögliche Offenleserahmen zu identifizieren, wurden die DNA-Sequenzen des gesamten H1.6- und eines Teils des E4.3-Fragments bestimmt. Die



Ergebnisse zeigten, daß, abgesehen von einem langen Streifen einer GC-reichen Sequenz, die die Erkennungsstellen von zwei Restriktionsenzymen (Bss HII und Sac II) trug, die sich häufig als mit untermethylierten CpG-Inseln in Zusammenhang stehend erwiesen, nur kurze Offenleserahmen vorhanden waren, die nicht ohne weiteres die starken Spezieskreuzungs-Hybridisationssignale erklären konnten.

Um den Methylierungsstatus dieser sehr CpC-reichen Region, die durch das Sequenzieren offenbart worden war, zu untersuchen, wurden genomische DNA-Sonden, die aus Fibroblasten und Lymphoblasten erzeugt worden waren, mit den Restriktionsenzymen Hpa II und Msp I digeriert und mittels Gelblothybridisation analysiert. Das Enzym Hpa II schneidet nur dann in die 5'-CCGG-3'-Sequenz, wenn das zweite Cytosin unmethyliert ist, wohingegen Msp I unabhängig vom Methylierungsstatus in diese Sequenz schneidet. Kleine DNA-Fragmente wurden von beiden Enzymen erzeugt, was darauf hinweist, daß diese CpG-reiche Region in der Tat in der genomischen DNA untermethyliert ist. Die Gelblothybridisation mit dem E4.3-Segment (Fig. 6) zeigt mit beiden Enzymen sehr kleine Hybridisationsfragmente, was auf die Anwesenheit einer hypomethylierten CpG-Insel hinweist.

Die vorstehend genannten Ergebnisse sind ein deutlicher Hinweis auf die Anwesenheit einer Kodierungsregion an diesem Locus. Zwei DNA-Segmente (E4.3 und H1.6), die Spezieskreuzungs-Hybridisationssignale aus diesem Bereich nachwiesen, wurden als Sonden verwendet, um cDNA-Bibliotheken, die aus mehreren Geweben und Zelltypen hergestellt worden waren, zu screenen.

cDNA-Bibliotheken aus kultivierten Epithelzellen wurden wie folgt hergestellt. Schweißdrüsenzellen, die aus einem nicht mit CF befallenen Individuum und von einem CF-Patienten erhalten worden waren, wurden wie bereits beschrieben [G. Collie et al., *In Vitro Cell. Dev. Biol.* 21, 592, 1985] zu einer ersten Passage gezüchtet. Die Anwesenheit von nach außen gerichteten Kanälen wurde in diesen Zellen bestätigt (J.A. Tabcharani, T.J. Jensen, J.R. Riordan, J.W. Hanrahan, *J. Memb. Biol.*, in Druck), aber die CF-Zellen waren gegenüber der Aktivierung durch cyclische AMP unempfind-



lich (T.J. Jensen, J.W. Hanrahan, J.A. Tabcharani, M. Buchwald und J.R. Riordan, *Pediatric Pulmonology*, Ergänzungsband 2, 100, 1988). RNA wurde aus ihnen nach dem Verfahren von J.M. Chirgwin et al. isoliert (*Biochemistry* 18, 5294, 1979). Poly A+RNA wurde selektiert (H. Aviv und P. Leder, *Proc. Natl. Acad. Sci. USA* 69, 1408, 5 1972) und als Schablone für die Synthese von cDNA mit Oligo (dT) -12-18 als Primer verwendet. Der zweite Strang wurde gemäß Gubler und Hoffman (*Gene* 25, 263, 1983) synthetisiert. Dieser wurde mit Eco RI-Methylase methyliert und seine Enden wurden mit T4-DNA-Polymerase geglättet. Phosphorylierte Eco RI-Linker wurden mit der cDNA ligiert und mit Eco RI begrenzt. Die Entfernung der überschüssigen Linker und 10 eine teilweise Größenfraktionierung wurde durch Chromatographie mit Biogel A-50 erreicht. Dann wurden die cDNAs in die Eco RI-Stelle der im Handel erhältlichen lambda ZAP ligiert. Rekombinanten wurden in *E. coli* BB4 gepackt und vermehrt. Teile der Packungsmischungen wurden amplifiziert und der Rest wurde für das Screenen vor der Amplifizierung zurückbehalten. Genauso wurde vorgegangen, um eine Bibliothek 15 aus RNA zu konstruieren, die aus prekonfluenten Kulturen der T-84 Dickdarm-Carcinoma-Zelllinie isoliert worden war (Dharmasathaphorn, K. et al., *Am. J. Physiol.* 246, G204, 1984). Die Zahl der unabhängigen Rekombinanten in den drei Bibliotheken betrug: 2×10^6 für die nicht mit CF befallenen Schweißdrüsenzellen, $4,5 \times 10^6$ für die CF-Schweißdrüsenzellen und $3,2 \times 10^6$ für T-84-Zellen. Diese Phagen wurden zu 50000 20 pro 15 cm Platte aufgestrichen und es wurden Plaque-Lifts unter Verwendung von Nylonmembranen (Biodyne) hergestellt und mit DNA-Fragmenten, die mit ^{32}p markiert waren, unter Verwendung der DNA-Polymerase I und einer Zufallsmischung aus Oligonucleotiden als Primer sondiert. Die Hybridisationsbedingungen waren gemäß G.M. Wahl und S.L. Berger (*Meth. Enzymol.* 152, 415, 1987). Bluescript™-Plasmide 25 wurden aus von Plaque gereinigten Klonen durch Excision mit einem M13-Helferphagen gewonnen. Die Lungen- und Bauchspeicheldrüsen-Bibliotheken wurden von Clontech Lab Inc. bezogen und hatten angegebene Größen von $1,4 \times 10^6$ und $1,7 \times 10^6$ unabhängigen Klonen.

30 Nach dem Screening von 7 verschiedenen Bibliotheken, von denen jede 1×10^5 bis 5×10^6 unabhängige Klone enthielt, wurde 1 einzelner Klon (als 10-1 identifiziert) mit



H1.6 aus einer cDNA-Bibliothek isoliert, die aus den kultivierten Schweißdrüsen-Epithelzellen eines nicht (mit CF) befallenen Individuums erzeugt worden war.

Die DNA-Sequenzanalyse zeigte, daß die Sonde 10-1 ein Insert mit einer Größe
5 von 920 bp und einen potentiellen, langen Offenleserahmen (ORF) enthielt. Da ein
Ende der Sequenz genau die gleiche Sequenzidentität hatte wie H1.6, wurde ge-
schlossen, daß der cDNA-Klon wahrscheinlich aus dieser Region stammte. Die gemein-
same DNA-Sequenz war jedoch nur 113 bp lang (siehe Fig. 4 und 7). Wie nachstehend
detailliert dargestellt, entsprach diese Sequenz tatsächlich dem 5' am nächsten liegenden
10 Exon auf dem mutmaßlichen CF-Gen. Der kurze Sequenz-Überlappungsbereich erklärte
somit die schwachen Hybridisationssignale beim Bibliotheken-Screening und die Un-
möglichkeit, Transkripte bei der RNA-Gelblotanalyse zu entdecken. Darüber hinaus
wurde die Orientierung der Transkriptionseinheit aufgrund der vergleichenden An-
ordnung der genomischen DNA-Sequenz mit dem angenommenen ORF von 10-1 vor-
15 läufig festgelegt.

Da man aufgrund von RNA-Gelblot-Hybridisationsversuchen annahm, daß das
entsprechende Transkript eine Länge von ungefähr 6500 Nucleotiden aufweist, war ein
weiteres cDNA-Bibliotheken-Screening erforderlich, um den Rest der Kodierungsregion
20 zu klonieren. Als Ergebnis von mehreren aufeinanderfolgenden Screenings mit cDNA-
Bibliotheken, die aus der Dickdarm-Carcinoma-Zelllinie T84, normalen und CF-
Schweißdrüsenzellen, Pancreas und adulten Lungen erzeugt worden waren, wurden 18
zusätzliche Klone isoliert (Fig. 7, wie im Anschluß in größerer Einzelheit erörtert). Die
DNA-Sequenzanalyse zeigte, daß keiner dieser cDNA-Klone der Länge des betrachteten
25 Transkripts entsprach, es war jedoch möglich, aufgrund von sich überlagernden
Regionen eine Consensus-Sequenz abzuleiten. Zusätzliche cDNA-Klone, die den 5'-
und 3'-Enden des Transkripts entsprachen, wurden aus 5'- und 3'-Primerextensions-
versuchen abgeleitet. Zusammen umfassen diese Klone insgesamt etwa 6,1 kb und
enthalten einen ORF, der in der Lage ist, für ein Polypeptid aus 1480 Aminosäureresten
30 zu kodieren (Fig. 1).



Es war ungewöhnlich zu beobachten, daß die meisten der hier isolierten cDNA-Klone Sequenzinsertionen an verschiedenen Stellen der Restriktionskarte von Fig. 7 enthielten. Die Karte zeigt Details der Genomstruktur des CF-Gens. Die Exon/Intron-Grenzen sind da angegeben, wo alle isolierten cDNA-Klone schematisch in der oberen Hälfte der Figur dargestellt sind. Viele dieser Extrasequenzen stimmen klar mit Intronregionen überein, die während der Konstruktion der cDNA umgekehrt transkribiert worden waren, wie sich bei der vergleichenden Anordnung mit genomischen DNA-Sequenzen zeigt.

Da die Zahl der rekombinanten cDNA-Klone für das CF-Gen, die beim Bibliotheken-Screening entdeckt wurden, viel kleiner war als aufgrund der Fülle an Transkripten, die nach RNA-Hybridisierungsversuchen geschätzt worden war, zu erwarten gewesen wäre, schien es wahrscheinlich, daß vorzugsweise die Klone, die abweichende Strukturen enthielten, zurückgeblieben waren, während die korrekten Klone während der Vermehrung verloren gegangen waren. Übereinstimmend mit dieser Interpretation wurde ein schlechtes Wachstum der Mehrheit der rekombinanten Klone, die in dieser Studie isoliert worden waren, beobachtet, unabhängig vom verwendeten Vektor.

Die Vorgehensweisen, um die 5'- und 3'-Enden der cDNA zu erhalten, waren denen ähnlich, die bereits beschrieben worden sind (M. Frohman et al., *Proc. Nat. Acad. Sci.*, USA, 85, 8998-9002, 1988). Für die 5'-Endklone wurden vollständige Pancreas- und T84 Poly A + RNA-Sonden umgekehrt transkribiert, wobei ein Primer (10b) verwendet wurde, der für Exon 2 spezifisch ist, ähnlich wie für die Primer-Extensionsreaktion beschrieben worden ist, außer daß ein radioaktiver Tracer an der Reaktion beteiligt wurde. Die Fraktionen, die bei der Synthese des ersten Strangs aus einer Agaroseperlsäule gesammelt wurden, wurden die durch Polymerase-Chainreaction (PCR) von eluierten Fraktionen analysiert. Die verwendeten Oligonucleotide lagen innerhalb der 10-1-Sequenz (145 Nucleotide auseinander). Aus den ersten Fraktionen, die ein PCR-Produkt lieferten, wurde ein Pool gebildet, und sie wurden durch Verdampfung aufkonzentriert und anschließend mit terminaler Desoxynucleotidyl-Transferase (BRL Labs) und dATP, wie vom Hersteller (BRL Labs) empfohlen, mit einem



Schwanz versehen. Dann wurde eine zweite Strangsynthese mittels eines Oligonucleotids, das eine abgeschlossene Linkersequenz 5'CGGAATTCTCGAGATC(T)₁₂3' enthielt, mit Taq-Polymerase (Cetus, AmpliTaq™) durchgeführt.

5 Dann wurde eine Amplifizierung mittels eines Anker (PCR) -Versuchs unter Verwendung der Linkersequenz und eines Primers, der gerade innerhalb des Extensionsprimers lag, der die Eco RI-Restriktionsstelle an seinem 5'-Ende besaß, durchgeführt. Im Anschluß an die Restriktion mit den Enzymen Eco RI und Bgl II und die Agarosegel-Reinigung wurden nach der Größe selektierte Produkte in das Plasmid Bluescript
10 KS, erhältlich von Stratagene, mittels Standardverfahren kloniert (Maniatis et al., s.o.). Praktisch alle gewonnenen Klone enthielten Inserte von weniger als 350 Nucleotiden. Um die 3'-Enden-Klone zu erhalten, wurde zuerst ein cDNA-Strang mit umgekehrter Transkription von 2 µg T84 Poly A + RNA unter Verwendung des mit einem Schwanz versehenen Linker-Oligonucleotids, das zuvor beschrieben wurde, mit Bedingungen, die
15 denen der Primerextension ähnlich sind, hergestellt. Dann wurde die Amplifizierung durch PCR mit dem Linker-Oligonucleotid und drei verschiedenen Oligonucleotiden, die bekannten Sequenzen des Klons T16-4.5 entsprachen, durchgeführt. Mit einem dieser Nucleotide mit der Sequenz 5'-ATGAAGTCCAAGGATTTAG-3' wurde eine Reaktion im präparativen Maßstab (2 x 100 µl) durchgeführt.

20 Dieses Oligonucleotid befindet sich ungefähr 70 Nucleotide flüßaufwärts von einer Hind III-Stelle innerhalb der bekannten Sequenz von T16-4.5. Der Restriktion des PCR-Produkts mit Hind III und Xho 1 folgte eine Agarosegel-Reinigung, um eine Bande bei 1,0 – 1,4 kb nach der Größe zu selektieren. Dieses Produkt wurde dann in das
25 Plasmid Bluescript KS kloniert, das von Stratagene erhältlich ist. Ungefähr 20 % der erhaltenen Klone hybridisierten an den 3'-Endabschnitt von T16-4.5. 10/10 der aus diesen Klonen isolierten Plasmide wiesen identische Restriktionskarten mit Insertgrößen von ungefähr 1,2 kb auf. Alle PCR-Reaktionen wurden nach der Empfehlung eines Enzymherstellers 30 Durchgänge lang in einem Puffer durchgeführt.

30



Ein Extensionsprimer, der 157 nt vom 5'-Ende des 10-1-Klons entfernt war, wurde verwendet, um den Startpunkt des vermuteten CF-Transkripts zu identifizieren. Der Primer wurde mit γ [³²p]ATP bei 5000 Curies/Mol und T4-Polynucleotidkinase endmarkiert und durch Spuncolumn-Gelfiltration gereinigt. Der radiomarkierte Primer wurde dann mit 4 – 5 μ g Poly A + RNA, hergestellt aus T-84 Dickdarm-Carcinoma-

5 zellen, in umgekehrtem 2X Transkriptase-Puffer 2 Stunden lang bei 60°C erhitzt. Im Anschluß an die Verdünnung und Zugabe von umgekehrter AMV-Transkriptase (Life Sciences, Inc.) wurde die Inkubation bei 41°C 1 Stunde lang fortgesetzt. Die Probe wurde dann mit 0,4M NaOH und 20 mM EDTA reguliert und schließlich mit NH₄OAc,

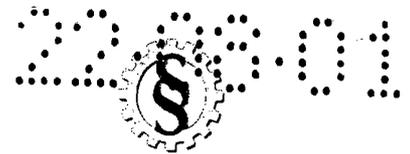
10 pH 4.6, neutralisiert, mit Phenol extrahiert, mit Ethanol ausgefällt, erneut in Puffer mit Formamid gelöst und auf einem Polyacrylamid-Sequenzierungsgel analysiert. Details dieser Verfahren wurden bereits beschrieben (*Meth. Enzymol.* 152, 1987, Herausg. S.L. Berger, A.R. Kimmel, Academic Press, N.Y.).

15 Die Ergebnisse des Primerextensionsversuchs unter Verwendung eines Extensions-Oligonucleotidprimers, bei dem 157 Nucleotide vom 5'-Ende von 10-1 ausgingen, sind in Tafel A von Fig. 10 gezeigt. Ein endmarkierter ØX174 Bakteriophage, der mit Hae II (BRL Labs) digeriert wurde, wird als Größenmarker verwendet. Zwei Hauptprodukte wurden nach 216 und 100 Nucleotiden beobachtet. Die Sequenz, die 100

20 Nucleotiden im 10-1 entspricht, entspricht einer sehr GC-reichen Sequenz (11/12), was nahelegt, daß es sich dabei um eine Pausenstelle des umgekehrten Transkripts handeln könnte. Die 5'-Anker-PCR-Ergebnisse sind in Tafel B von Fig. 10 gezeigt. Das 1,4%-ige Agarosegel, das links gezeigt ist, wurde auf Zetaprobe™-Membran (Bio-Rad Lab.) geblottet und transferiert. Die DNA-Gelblot-Hybridisation mit radiomarkiertem

25 10-1 ist rechts gezeigt. Man sieht, daß die Größe der 5'-Extensionsprodukte von 170 bis 280 nt variiert, wobei das Hauptprodukt etwa 200 Nucleotide zeigt. Die PCR-Kontrollstraße zeigt ein Fragment aus 145 Nucleotiden. Es wurde durch Verwendung der Testoligomere innerhalb der 10-1 Sequenz erhalten. Die gezeigten Größenmarker entsprechen den Größen 154, 220/210, 298, 344, 394 Nucleotide (1kb-Leiter erworben von

30 BRL Lab).



Das unter Tafel B von Fig. 10 gezeigte Schema schildert die Vorgehensweise zum Erhalt der doppelsträngigen cDNA, die für die Amplifizierung und Klonierung befolgt wurde, um die Klone PA3-5 und TB2-7, die in Fig. 7 gezeigt sind, zu erzeugen. Die PCR-Ankerversuche zur Beschreibung des 3'-Endes sind in Tafel C gezeigt. Wie im Schema unter Fig. 10C dargestellt, wurden drei Primer, deren Position zueinander bekannt war, für die Amplifizierung mit umgekehrter transkribierter T84 RNA, wie beschrieben, verwendet. Diese Produkte wurden auf einem 1%-igen Agarosegel separiert und auf Nylonmembran wie vorstehend beschrieben geblottet. Die DNA-Blothybridisation mit dem 3'-Abschnitt des T16-4.5-Klons ergab Banden mit Größen, die dem Abstand zwischen dem verwendeten spezifischen Oligomer und dem 3'-Ende des Transkripts entsprachen. Diese Banden in den Straßen 1, 2a und 3 sind schematisch unter Tafel C in Fig. 10 gezeigt. Die Bande in Straße 3 ist schwach, da nur 60 Nucleotide dieses Segments die verwendete Sonde überlappen. Ebenfalls im Schema angezeigt und wie in Straße 2b gezeigt ist das Produkt, das durch Restriktion des PCR-Ankerprodukts erzeugt wurde, um das Klonieren zu erleichtern, mit dem der THZ-4-Klon erzeugt wird, der in Fig. 7 gezeigt ist.

Die DNA-Blothybridisationsanalyse der mit EcoRI und HindIII-Enzymen digestierten genomischen DNA, die mit Abschnitten der cDNAs, die das gesamte Transkript umfassen, digeriert wurde, legt nahe, daß das Gen mindestens 24 Exons enthält, die mit den römischen Zahlen I bis XXIV numeriert sind (siehe Fig. 9). Diese entsprechen den Zahlen 1 bis 24, die in Fig. 7 gezeigt sind. Die Größe jeder Bande ist in kb angegeben. Wie mit Bezug auf Fig. 18 erörtert werden wird, wurde jedoch nachgewiesen, daß die Exons 6, 14 und 17 zweiteilig sind, und sie wurden als Exons 6a, 6b, 14a, 14b, 17a und 17b identifiziert, wodurch sich insgesamt 27 Exons im CFTR-Gen ergaben.

In der Fig. 7 zeigen leere Kästchen die ungefähren Positionen der 24 Exons an, die durch Isolation der > 22 Klone aus dem Screening von cDNA-Bibliotheken und aus den PCR-Ankerversuchen erhalten worden waren, die ausgelegt waren, die 5'- und 3'-Enden zu klonieren. Die Längen der genomischen Eco RI-Fragmente in kb, die von jedem Exon nachgewiesen werden, ist ebenfalls angegeben. Die schraffierten Kästchen in



Fig. 7 zeigen die Anwesenheit von Intronsequenzen an und die gepunkteten Kästchen zeigen andere Sequenzen an. Links unten von dem gefüllten Kästchen ist die relative Position des Klons H1.6 dargestellt, der verwendet wird, um den ersten cDNA-Klon 10-1 aus 10^6 Phagen der normalen Schweißdrüsen-Bibliothek nachzuweisen. Wie in Fig. 4(D) und 7 gezeigt, überlappt der genomische Klon H1.6 teilweise ein EcoRI-Fragment von 4,3 kb. Alle gezeigten cDNA-Klone wurden zu genomischer DNA hybridisiert und/oder wurden Restriktions-feinkartiert. Beispiele für die Restriktionsstellen, die innerhalb der cDNAs und in den entsprechenden Genomfragmenten auftreten, sind angegeben.

10

Fig. 9: die Hybridisationsanalyse schließt Sonden ein; d.h. cDNA-Klone 10-1 für Tafel A, T16-1 (3'-Abschnitt) für Tafel B, T16-4.5 (Mittelabschnitt) für Tafel C und T16-4.5 (3'-Endabschnitt) für Tafel D. In Tafel A der Fig. 9 weist die cDNA-Sonde 10-1 die genomischen Banden der Exons I bis VI nach. Der 3'-Abschnitt von T16-1, erzeugt durch NruI-Restriktion, weist die Exons IV bis XIII nach, wie in Tafel B gezeigt. Diese Sonde überlappt teilweise 10-1. Die Tafeln C bzw. D zeigen genomischen Banden, die von den mittleren und den 3'-End-EcoRI-Fragmenten des Klons T16-4.5 nachgewiesen wurden. Zwei EcoRI-Stellen treten innerhalb der cDNA-Sequenz auf und spalten die Exons XIII und XIX. Wie von den Exons in Klammern angezeigt, entsprechen jedem dieser Exons zwei genomische EcoRI-Banden. Es wurde eine Kreuzungshybridisation an andere genomische Fragmente beobachtet. Diese Banden, mit N bezeichnet, stammen nicht vom Chromosom 7, da sie nicht in Mensch/Hamster-Hybriden erschienen, die das menschliche Chromosom 7 enthielten. Man nimmt an, daß die schwache Bande in Tafel D, die durch XI in Klammern angezeigt ist, aufgrund einer internen Homologie mit der cDNA die Kreuzungshybridisation von Sequenzen verursacht.

Da 10-1 eine starke Bande bei der Gelblot-Hybridisation von RNA aus der T-84 Dickdarm Carcinoma-Zelllinie nachwies, wurde diese cDNA verwendet, um die Bibliothek, die mit dieser Quelle konstruiert wurde, zu screenen. Es wurden fünfzehn Positive erhalten, unter denen die Klone T6, T6/20, T11, T16-1 und T13-1 gereinigt und

30



sequenziert wurden. Ein erneutes Screening der gleichen Bibliothek mit einem 0,75 kb Bam HI Eco RI-Fragment vom 3'-Ende von T16-4.5 ergab T16-4.5. Ein 1,8 kb EcoRI-Fragment vom 3'-Ende von T16-4.5 ergab T8-B3 und T12a, wobei letzteres ein Polyadenylationssignal und einen Schwanz enthielt. Gleichzeitig wurde eine menschliche
5 Lungen-cDNA-Bibliothek gescreent; es wurden viele Klone isoliert, einschließlich derer, die hier mit der Vorsilbe 'CDL' angegeben sind. Es wurde auch eine Pancreas-Bibliothek gescreent, was den Klon CDPJ5 ergab.

Um Kopien dieses Transkripts aus einem CF-Patienten zu erhalten, wurde eine
10 cDNA-Bibliothek aus RNA von Schweißdrüsen-Epithelzellen eines Patienten mit dem 0,75 kb Bam HI – Eco RI-Fragment vom 3'-Ende von T16-1 gescreent und die Klone C16-1 und C1-1/5, die alles außer Exon 1 abdeckten, wurden isoliert. Diese beiden Klone zeigen beide eine 3 bp-Deletion in Exon 10, die in keinem anderen Klon vorkommt, der dieses Exon enthält. Mehrere Klone, einschließlich CDLS26-1 aus der
15 Lungenbibliothek und T6/20 und T13-1, isoliert von T84, wurden aus teilweise gereiften Transkripten erhalten. Dies wurde durch genomische Hybridisation und durch Sequenzierung über die Exon/Intron-Grenzen hinweg für jeden Klon bestätigt. T11 enthielt an jedem Ende auch zusätzliche Sequenz. T16-4.5 enthielt eine kleine Insertion in der Nähe der Grenze zwischen den Exons 10 und 11, die nicht der Intronsequenz
20 entsprach. Die Klone CDLS16A, 12a und 13a aus der Lungenbibliothek enthielten auch Fremdsequenzen unbekanntes Ursprungs. Der Klon C16-1 enthielt auch eine kurze Insertion, die einem Abschnitt des γ -Transposon von *E. coli* entsprach; dieses Element wurde in den anderen Klonen nicht nachgewiesen. Die 5'-Klone PA3-5, erzeugt aus Pancreas-RNA und TB2-7, hergestellt aus T84-RNA unter Verwendung der PCR-
25 Ankertechnik, weisen identische Sequenzen auf, abgesehen von einer einzigen Nucleotid-Längendifferenz in der am 5'-Ende, wie in Fig. 1 gezeigt. Der 3'-Klon, THZ-4, erhalten aus T84-RNA, enthält die 3'-Sequenz des Transkripts in Übereinstimmung mit der Genomsequenz dieser Region.

30 Eine kombinierte Sequenz, die die vermutete Kodierungsregion des CF-Gens darstellt, wurde aus sich überlagernden cDNA-Klonen erzeugt. Da die meisten der cDNA-



Klone offensichtlich aus unreifen Transkripts stammten, wurden weitere Studien durchgeführt, um die Authentizität der kombinierten Sequenz sicherzustellen. Jeder cDNA-Klon wurde zuerst bezüglich seiner Lage auf Chromosom 7 durch Hybridisationsanalyse mit Mensch/Hamster-Somazellhybrid, das ein einziges menschliches Chromosom 7 enthielt, und durch Pulsfeld-Gelelektrophorese getestet. Eine Restriktionsenzym-Feinkartierung wurde ebenfalls mit jedem Klon durchgeführt. Während sich überlagernde Regionen bei den meisten der Klone klar zu identifizieren waren, enthielten viele von ihnen Regionen mit einzigartigen Restriktionsmustern.

Um diese cDNA-Klone weiter zu beschreiben, wurden sie als Sonden in Gelhybridisationsversuchen mit EcoRI- oder HindIII-digierter genomischer menschlicher DNA verwendet. Wie in Fig. 9 gezeigt, konnten fünf bis sechs verschiedene Restriktionsfragmente mit den 10-1-cDNA- und eine ähnliche Anzahl von Fragmenten mit anderen cDNA-Klonen nachgewiesen werden, was auf das Vorhandensein von mehreren Exons für das vermutete CF-Gen schließen läßt. Bei den Hybridisationsstudien wurden auch die cDNA-Klone mit unreifen Intronsequenzen identifiziert, da sie bevorzugt eine Hybridisation an eine Untergruppe von genomischen DNA-Fragmenten zeigten. Von den nachgewiesenen cDNA-Klone wurden die entsprechenden genomischen DNA-Segmente isoliert und die Exons und die Exon/Intron-Grenzen wurden sequenziert. Wie in Fig. 7 angezeigt, wurden mindestens 27 Exons identifiziert, einschließlich der Spaltexons 6a, 6b, 14a, 14b und 17a, 17b. Aufgrund dieser Informationen und der Ergebnisse der physischen Kartierungsversuche wurde geschätzt, daß der Genlocus 250 kb auf Chromosom 7 umfaßt.

2.6 DIE SEQUENZ

Fig. 1 zeigt die Nucleotidsequenz der klonierten cDNA, die für CFTR kodiert, zusammen mit der abgeleiteten Aminosäuresequenz. Die erste Basenposition entspricht dem ersten Nucleotid im 5'-Extensionsklon PA3-5, der um ein Nucleotid länger ist als TB2-7. Die Pfeile zeigen die Position der Transkriptionsinitiationsstelle gemäß Primerextensionanalyse an. Auf das Nucleotid 6129 folgt ein Poly(dA) Trakt. Die Positionen



der Exon-Verbindungen sind durch vertikale Linien angezeigt. Potentielle Membran-
umspannende Segmente wurden mittels des Algorithmus von Eisenberg et al., *J. Mol.*
Biol. 179:125 (1984) bestätigt. Potentielle Membran-umspannende Segmente, wie ana-
lysiert und in Fig. 11 gezeigt, sind in den Kästchen der Fig. 1 enthalten. In Fig. 11 wird
5 der mittlere Hydropathie-Index [Kyte und Doolittle, *J. Molec. Biol.* 157:105 (1982)] von
9 Restpeptiden gegen die Aminosäurezahl aufgezeichnet. Die entsprechenden Positio-
nen der Merkmale der sekundären Struktur, die gemäß Garnier et al. [*J. Molec. Biol.*,
157, 165 (1982)]]vorausgesagt wurden, sind in der unteren Tafel angegeben. Amino-
säuren, die die vermuteten Bindungsfalten enthalten, sind in Fig. 1 unterstrichen.
10 Mögliche Stellen der Phosphorylisation durch die Proteinkinasen A (PKA) oder C
(PKC) sind durch offene bzw. geschlossene Kreise angegeben. Das offene Dreieck
befindet sich über den 3 bp (CTT), die in CF deletiert sind (siehe die nachstehende
Erörterung). Die cDNA-Klone in Fig. 1 wurden durch das Didesoxy-Kettenabbruchs-
verfahren mittels des automatischen DuPont Genesis 2000™-DNA-Sequenzierers, bei
15 denen ³⁵S-markierte Nucleotide verwendet wurden, sequenziert.

Die kombinierte cDNA-Sequenz umfaßt 6129 Basenpaare, ausschließlich des
Poly(A)-Schwanzes am Ende der nicht-translatierten 3'-Region, und enthält einen ORF,
der fähig ist, für ein Polypeptid aus 1480 Aminosäuren zu kodieren (Fig. 1). Ein ATG
20 (AUG)-Triplet ist am Anfang dieses ORF vorhanden (Basenposition 133 – 135). Da die
Nucleotidsequenz in der Umgebung dieses Codons (5'-AGACCAAUGCA-3') die ange-
nommenen Merkmale der Consensus-Sequenz (CC) A/GCCAAUGG(G) einer eukaryon-
tischen Translationsinitiationsstelle mit einem stark konservierten A an der -3-Position
aufweist, ist es sehr wahrscheinlich, daß dieses AUG dem ersten Methionincodon des
25 vermuteten Polypeptids entspricht.

Um die Sequenz zu erhalten, die dem 5'-Ende des Transkripts entspricht, wurde
ein Primerextensionsversuch durchgeführt, wie an früherer Stelle beschrieben. Wie in
Fig. 10A gezeigt, konnte ein Primerextensionsprodukt aus ungefähr 216 Nucleotiden
30 beobachtet werden, was darauf hindeutet, daß das 5'-Ende des Transkripts ungefähr 60
Nucleotide flußaufwärts vom Ende des cDNA-Klons 10-1 initiierte. Eine modifizierte



Polymerase-Chainreaction (Anker-PCR) wurde dann verwendet, um das Klonieren der 5'-Endsequenzen zu erleichtern (Fig. 10b). Zwei unabhängige 5'-Extensionsklone, einer aus Pancreas- und der andere aus T84-RNA, wurden durch DNA-Sequenzieren beschrieben und man fand, daß sie sich nur in der Länge 1 Base unterschieden, was die
5 wahrscheinlichste Initiationsstelle für das Transkript anzeigte, wie in Fig. 1 gezeigt.

Da die meisten der cDNA-Initialklone keinen polyA-Schwanz enthielten, der das Ende einer mRNA anzeigt, wurde auch mit dem 3'-Ende des Transkripts eine Anker-PCR durchgeführt (Frohman et al., 1988, s.o.). Drei 3'-Extensionsoligonucleotide
10 wurden zum Endabschnitt des cDNA-Klons T16-4.5 gemacht. Wie in Fig. 10c gezeigt, wurden 3 PCR-Produkte mit unterschiedlicher Größe erhalten. Alle stimmten mit der Interpretation überein, daß das Ende des Transkripts sich ungefähr 1,2 kb flußabwärts von der HindIII-Stelle an der Nucleotidposition 5027 befand (siehe Fig. 1). Die DNA-Sequenz, die aus repräsentativen Klonen erhalten wurde, stimmte mit der des T84-
15 cDNA-Klons T12a (siehe Fig. 1 und 7) und der Sequenz des entsprechenden 2,3 kb EcoRI-Genomfragments überein.

3.0 MOLEKULARGENETIK VON CF

3.1 EXPRESSIONSSTELLEN

20

Um das Transkript für das vermutete CF-Gen zu visualisieren, wurden RNA-Gel-blot-Hybridisationsversuche mit der 10-1-cDNA als Sonde durchgeführt. Die Ergebnisse der RNA-Hybridisation sind in Fig. 8 gezeigt.

25

Die RNA-Proben wurden gemäß Verfahren, die bereits beschrieben worden sind (A.M. Kimmel, S.L. Berger, Herausg. *Meth. Enzymol.* 152, 1987) aus Gewebeproben hergestellt, die aus der chirurgischen Pathologie oder bei einer Autopsie erhalten wurden. Formaldehydgele wurden auf Nylonmembrane transferiert (Zetaprobe™; BioRad Lab). Die Membrane wurden dann gemäß bereits veröffentlichten Verfahren
30 (J. Rommens et al., *Am. J. Hum. Genet.* 43, 645-663, 1988) mit DNA-Sonden hybridisiert, die durch das Zufallsprimingverfahren auf hohe spezifische Aktivität markiert



worden waren (A.P. Feinberg und B. Vogelstein, *Anal. Biochem.* 132, 6, 1983). Fig. 8 zeigt die Hybridisation durch den cDNA-Klon 10-1 auf einem 6,5 kb Transkript in den angegebenen Geweben. Vollständige RNA (10 µg) jedes Gewebes und Poly A+ RNA (1 µg) der T84 Dickdarm-Carcinoma-Zelllinie wurden auf einem 1%-igen Formaldehydgel separiert. Die Positionen der 28S- und 18S-rRNA-Banden sind angegeben. Die Pfeile zeigen die Position des Transkripts an. Die Größenbestimmung wurde durch Vergleich mit Standard-RNA-Markern durchgeführt (BRL Labs). HL60 ist eine menschliche promyelocytische Leukämiezelllinie, und T84 ist eine menschliche Dickdarmkrebs-Zelllinie.

Die Analyse zeigt eine auffällige Bande mit einer Größe von ungefähr 6,5 kb in T84-Zellen. Ähnlich dazu wurden starke Hybridisationssignale auch in Pancreas- und primären Kulturen von Zellen aus Nasenpolypen nachgewiesen, was darauf hinweist, daß die reife mRNA des vermuteten CF-Gens ungefähr 6,5 kb umfaßt. Schwächere Hybridisationssignale, die vermutlich Abbauprodukte darstellen, wurden in den geringeren Größenbereichen nachgewiesen, sie variierten jedoch von Versuch zu Versuch. Identische Resultate wurden mit anderen cDNA-Klonen als Sonden erhalten. Aufgrund der Hybridisations-Bandenintensität und des Vergleichs mit denen, die für andere Transkripte unter identischen Versuchsbedingungen nachgewiesen worden waren, wurde geschätzt, daß die vermuteten CF-Transkripts ungefähr 0,01 % der gesamten mRNA in T84-Zellen ausmachten.

Eine Reihe anderer Gewebe wurden in einem Versuch, um das Expressionsmuster des 10-1-Gens und die CF-Pathologie in Zusammenhang zu bringen, auch durch RNA-Gelblot-Hybridisationsanalyse untersucht. Wie in Fig. 8 gezeigt, wurden Transkripte, die alle gleich groß waren, in der Lunge, dem Dickdarm, den Schweißdrüsen (kultivierte Epithelzellen), der Plazenta, der Leber und der Ohrspeicheldrüse gefunden, aber die Signalintensitäten dieser Gewebe variierten von Präparation zu Präparation und waren im allgemeinen schwächer als diejenigen, die im Pancreas oder in Nasenpolypen gefunden wurden. Die Intensität variierte von Präparation zu Präparation, beispielsweise wurde in der Präparation, die in Fig. 8 gezeigt ist, keine Hybridisation in der Niere



entdeckt, konnte aber in anschließenden wiederholten Assays wahrgenommen werden. Im Gehirn oder der Nebenniere konnten keine Hybridisationssignale wahrgenommen werden (Fig. 8), auch nicht in Hautfibroblast- und Lymphoblast-Zelllinien.

5 Insgesamt schien die Expression des CF-Gens in vielen der untersuchten Gewebe aufzutreten, wobei in den Geweben, die ernsthaft von CF befallen sind, höhere Niveaus auftraten. Obwohl dieses Epithelgewebe-spezifische Expressionsmuster gut mit der Pathologie der Krankheit übereinstimmt, wurde kein deutlicher Unterschied in der Menge oder der Größe der Transkripts von CF- und Kontrollgeweben nachgewiesen,
10 was die Annahme bestätigt, daß CF-Mutationen subtile Veränderungen auf Nucleotid-Niveau darstellen.

3.2 DIE HAUPT-CF-MUTATION

15 Fig. 16 zeigt die DNA-Sequenz an der F508-Deletion. Links das umgekehrte Komplement der Sequenz von der Basenposition 1649-1664 der normalen Sequenz (wie vom cDNA-Klon T16 abgeleitet). Die Nucleotidsequenz ist für jede der 2 Signalverstärkerröhren (PMT Nr. 1 und Nr. 2) eines DuPont Genesis 2000™-DNA-Analysesystems als Ausgangssignal (in ungerichteten Fluoreszenz-Intensitätseinheiten,
20 y-Achse), aufgezeichnet gegen die Zeit (x-Achse), dargestellt. Die entsprechende Nucleotidsequenz ist darunter gezeigt. Rechts ist die gleiche Region von einer Mutantensequenz (wie vom cDNA-Klon C16 erhalten). Doppelsträngige DNA-Plasmidschablonen wurden anhand des alkalischen Lyseverfahrens hergestellt. Fünf µg Plasmid-DNA und 75 ng Oligonucleotidprimer wurden in jeder Sequenzierungsreaktion
25 gemäß dem von DuPont empfohlenen Protokoll verwendet, abgesehen davon, daß das Erhitzen bei 45°C 30 min lang durchgeführt wurde und daß der Elongations/Terminations-Schritt 10 min lang bei 42°C durchgeführt wurde. Die nicht-inkorporierten fluoreszierenden Nucleotide wurden durch Ausfällung des DNA-Sequenzierungs-Reaktionsprodukts mit Ethanol in der Anwesenheit von 2,5 M Ammoniumacetat bei
30 einem pH von 7,0 entfernt und einmal mit 70 % Ethanol gespült. Der für das T16-1-Sequenzieren verwendete Primer war ein spezifisches Oligonucleotid



5'-GTTGGCATGCTTTGATGACGCTTC-3' und der für C16-1 war der Universal-primer SK für den Bluescript-Vektor (Stratagene).

Fig. 17 zeigt auch die DNA-Sequenz in der Umgebung der F508-Deletion, wie durch manuelles Sequenzieren bestimmt. Die normale Sequenz von der Basenposition 1645-1671 (von cDNA T16-1) ist neben der CF-Sequenz gezeigt (von cDNA C16-1). Die linke Tafel zeigt die Sequenzen der Kodierstränge, die mit dem B-Primer (5'-GTTTCCTGGATTATGCCTGGCAC-3') erhalten wurden, und die rechte Tafel zeigt diejenigen vom gegenüberliegenden Strang, die mit dem D-Primer (5'-GTTGGCATGCTTTGATGACGCTTC-3') erhalten wurden. Die Klammern zeigen die drei Nucleotide in der normalen Sequenz an, die in CF-Sequenz fehlen (Pfeilspitzen). Das Sequenzieren wurde durchgeführt wie in F. Sanger, S. Nicklen, A.R. Coulsen, *Proc. Nat. Acad. Sci. U.S.A* 74: 5463 (1977) beschrieben.

Die umfangreichen genetischen und physischen Kartierungsdaten haben den molekularen Klonierungsstudien eine Richtung gegeben, die sich auf ein kleines DNA-Segment auf dem Chromosom 7 konzentriert. Aufgrund des Fehlens von Chromosomendeletionen und -umgruppierungen in CF und dem Fehlen eines hochentwickelten funktionellen Assays für das CF-Genprodukt erforderte die Identifikation des CF-Gens eine detaillierte Beschreibung des Locus selbst und einen Vergleich zwischen den CF- und normalen (N) Allelen. Zufällig ausgewählte, phänotypisch normale Individuen konnten aufgrund der Häufigkeit von symptomfreien Trägern in der Bevölkerung nicht als Kontrollpersonen für den Vergleich einbezogen werden. Infolgedessen eigneten sich nur Eltern von CF-Patienten, von denen jeder *per definitionem* ein N- und ein CF-Chromosom trägt, für die Analyse. Außerdem war es aufgrund des starken allelischen Zusammenhangs, der zwischen CF und einigen der eng gekoppelten DNA-Markern beobachtet wurde, erforderlich, die Möglichkeit auszuschließen, daß es sich bei Sequenzunterschieden, die zwischen N und CF nachgewiesen wurden, um Polymorphismen handelte, die mit dem Krankheitslocus assoziiert waren.

30



3.3 IDENTIFIKATION VON RFLPs UND FAMILIENSTUDIEN

Um die Beziehung von jedem der DNA-Segmente, die aus den Chromosomen-Walking- und -Jumping-Versuchen zu CF isoliert worden waren, zu bestimmen, wurden Restriktionsfragmentlängen-Polymorphismen (RFLPs) identifiziert und verwendet, um Familien zu studieren, bei denen zuvor Crossingover-Vorgänge zwischen CF und anderen flankierenden DNA-Markern nachgewiesen worden waren. Wie in Fig. 14 gezeigt, wurden insgesamt 18 RFLPs in der 500 kb Region nachgewiesen; 17 davon (von E6 bis CE1.0) sind in Tabelle 2 aufgeführt; einige von ihnen entsprachen Markern, über die zuvor berichtet worden war.

Fünf der RFLPs, nämlich 10-1X.6, T6/20, H1.3 und CE1.0, wurden mit cDNA- und genomischen DNA-Sonden, die vom vermuteten CF-Gen abgeleitet worden waren, identifiziert. Die RFLP-Daten sind in Tabelle 2 angegeben, wobei Marker in der MET- und D7S8-Region aus Vergleichsgründen eingeschlossen wurden. Die physischen Entfernungen zwischen diesen Markern, ebenso wie ihre Beziehung zu den MET- und D7S8-Regionen, sind in Fig. 14 gezeigt.



TABELLE 2. MIT DEM CF-GEN ASSOZIIERTE RFLPs

Sondenname	Enzym	Fragmentlänge	N(a)	CF-PI(a)	A(b)	*(c)	Referenz
metD	BamI	7,6 (kb)	28	48	0,60	0,10	J.E. Spence et al., <i>Am. J. Hum. Genet.</i> 39:729 (1986)
		6,8	59	25			
metD	TaqI	6,2	74	75	0,66	0,06	R. White et al., <i>Nature</i> 318:382 (1985)
		4,8	19	4			
metH	TaqI	7,5	45	49	0,35	0,05	White et al., s.o.
		4,0	38	20			
E6	TaqI	4,4	58	62	0,45	0,06	B. Keren et al., <i>Am. J. Hum. Genet.</i> 44: 827 (1989)
		3,6	42	17			
E7	TaqI	3,9	40	16	0,47	0,07	
		3+0,9	51	57			
pH131	HinfII	0,4	81	33	0,73	0,15	J.M. Rommens et al., <i>Am. J. Hum. Genet.</i> 43:645 (1988)
		0,3	18	47			
W3D1.4	HindIII	20	82	33	0,68	0,13	B. Kerem et al., s.o.
		10	22	47			
H2.3A	TaqI	2,1	39	53	0,64	0,09	X. Estivill et al., <i>Nature</i> 326: 840 (1987); X. Estivill et al., <i>Genomics</i> 1: 257 (1987)





TABELLE 2 (Fortsetzung)

(XV2C)		1,4	37	11		
EG1.4	HincII	3,8	31	69	0,89	0,17
		2,8	56	7		
EG1.4	BgII	20	27	69	0,89	0,18
		15	62	9		
JG2E1	PstI	7,8	69	10	0,88	0,18
						X. Estivill et al., s.o., und B. Kerem et al., s.o., (KM19) 6.63070
E2.6/E.9	MspI	13	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
(Mp6d9)		8,5+3,5	38	64		G. Romeo, in persönlicher Kommunikation
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





Tabelle 2 (Fortsetzung)

T6/20	MspI	0,6	72	15	
		8	56	66	0,54
		4,3	21	8	
H1.3	NcoI	2,4	53	7	0,87 0,15
		1+1,4	35	69	
CE1.0	NdeI	5,5	81	73	0,41 0,03
		4,7+0,8	8	3	
J32	SacI	15	21	24	0,17 0,02 M.C. Ianuzzi et al., <i>Am. J. Genet.</i> 44: 695 (1989)
		6	47	38	
J3.11	MspI	4,2	36	38	0,29 0,04 B.J. Wainright et al., <i>Nature</i> 318: 384 (1985)
		1,8	62	36	
J29	PvuII	9	26	36	0,36 0,06 M.C. Ianuzzi et al., s.o.
		6	55	36	

0901



BEMERKUNGEN ZUR TABELLE 2

- 5 (a) Die Zahl der N- und CF-PI- (CF mit Bauchspeicheldrüsen-Unterfunktion) Chromosomen wurde von den Eltern der Familien abgeleitet, die bei der Kopplungsanalyse verwendet worden waren [Tsui et al., *Cold Spring Harbor Symp. Quant. Biol.* 51:325 8f1986)].
- 10 (b) Eine standardisierte Assoziation (A), die weniger durch die Fluktuation der DNA-Markerallelverteilung in den N-Chromosomen beeinflusst ist, wird hier für den Vergleich verwendet. Yule's Assoziationskoeffizient ($A = (ad - bc)/(ad + bc)$), wobei a, b, c und d die Zahl der N-Chromosomen mit dem DNA-Markerallel 1, CF mit 1, N mit 2 bzw. CF mit 2 darstellt. Das relative Risiko kann mittels der Beziehung $PR = (1 + A)/(1 - A)$ oder seiner Umkehrung berechnet werden.
- 15 (c) Allelische Assoziation (*), berechnet gemäß A. Chakravarti et al., *Am. J. Hum. Genet* 36:1239 (1984), die angenommene Häufigkeit von 0,02 CF-Chromosomen in der Bevölkerung ist für den Vergleich eingeschlossen.

Aufgrund der niedrigen Zahl von rekombinanten Familien, die für die Analyse zur Verfügung standen, was wegen der geringen Distanz zwischen den untersuchten Markern und CF zu erwarten gewesen war, und wegen der Möglichkeit von Fehldiagnosen waren alternative Ansätze für das weitere Feinkartieren des CF-Gens notwendig.

3.4 ALLELISCHE ASSOZIATION

25 Eine allelische Assoziation (Verbindungsungleichgewicht) wurde in vielen eng gekoppelten DNA-Markern nachgewiesen. Obwohl die Eignung der allelischen Assoziation für das Messen der genetischen Distanz ungewiss ist, wurde eine allgemeine Korrelation zwischen CF und den flankierenden DNA-Markern beobachtet. Eine starke Assoziation mit CF wurde für die näher gelegenen DNA-Marker, D7S23 und D7S122
30 bemerkt, wogegen eine geringe oder keine Assoziation bei den weiter entfernten Markern MET oder D7S8 nachgewiesen wurde (siehe Fig. 3E).



Wie in Tabelle 2 gezeigt, stieg der Assoziationsgrad zwischen DNA-Markern und CF (wie anhand des Yule'schen Assoziationskoeffizienten gemessen) von 0,35 für metH und 0,17 für J32 auf 0,91 für 10-1X.6 (nur CF-PI-Patientenfamilien wurden bei der Analyse eingesetzt, da sie genetisch homogener zu sein scheinen als CF-PS). Die Assoziationskoeffizienten schienen über die 300 kb von EG1.4 bis H1.3 ziemlich konstant zu sein; die Fluktuation, die an mehreren Stellen, am auffälligsten H2.3A, E4.1 und T6/20, nachgewiesen wurde, war wahrscheinlich auf Variationen in der allelischen Verteilung innerhalb der N-Chromosomen zurückzuführen (siehe Tabelle 2). Diese Daten sind daher konsistent mit dem Ergebnis der Studie rekombinanter Familien (siehe Fig. 14). Eine ähnliche Schlußfolgerung konnte auch durch Untersuchung der extendierten DNA-Makerhaplotypen gezogen werden, die mit den CF-Chromosomen assoziiert sind. (siehe unten). Jedoch erlaubt die starke allelische Assoziation, die über die große physische Distanz zwischen EG1.4 und H1.3 entdeckt worden war, keine weitere Feinkartierung des CF-Gens. Da J44 der letzte genomische DNA-Klon war, der durch Chromosomen-Walking und -Jumping isoliert worden war, bevor ein cDNA-Klon identifiziert wurde, veranlaßte uns die starke allelische Assoziation, die für das JG2E1-J44-Intervall nachgewiesen wurde, dazu, in diesem gesamten Intervall nach Kandidatengensequenzen zu suchen. Dabei ist es interessant, daß der höchste Grad an allelischer Assoziation tatsächlich zwischen CF und den 2 RFLPs, die von 10-1X.6 nachgewiesen worden waren, einer Region in der Nähe der CF-Hauptmutation, nachgewiesen wurde.

Tabelle 3 zeigt eine paarweise allelische Assoziation zwischen DNA-Markern, die eng mit CF gekoppelt waren. Die durchschnittliche Zahl der Chromosomen, die bei diesen Berechnungen verwendet wurde, war 75 – 80, und nur Chromosomen von CF-PI-Familien wurden bei der Indexierung von CF-Chromosomen verwendet. Ähnliche Ergebnisse wurden erzielt, wenn Yule's standardisierte Assoziation (A) angewendet wurde.



TABELLE 3

N Chromosomen

	metD	metH	E6	E7	pH131	D1.4	H2-3A	W3	EG1.4	JG2E1	E2.6	H2.8	E4.1	J44	10-1X.6	T6/20	H1.3	CE1.0	J32	J3.11	J29	
	BanI	TaqI	TaqI	TaqI	HinII	HdIII	TaqI	HcII	BglI	PstI	MspI	NcoI	MspI	XbaI	AccI	HaeIII	MspI	NcoI	NdeI	SacI	MspI	PvuII
metD BanI	0.35	0.49	0.04	0.04	0.05	0.07	0.27	0.08	0.08	0.07	0.14	0.07	0.09	0.03	0.06	0.10	0.03	0.16	0.05	0.07	0.11	0.02
metD TaqI	0.21	-	0.41	0.13	0.15	0.02	0.01	0.02	0.09	0.15	0.11	0.07	0.24	0.03	0.11	0.08	0.02	0.06	0.13	0.15	0.09	0.09
metH TaqI	0.81	0.14	-	0.01	0.05	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	-	0.93	0.07	0.08	0.04	0.02	0.03	0.00	0.19	0.02	0.09	0.11	0.09	0.15	0.07	0.11	0.20	0.00	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
pH131 HinII	0.45	0.28	0.23	0.38	0.40	-	0.91	0.12	0.04	0.09	0.05	0.06	0.03	0.03	0.10	0.12	0.10	0.23	0.10	0.05	0.05	0.10
W3D1.4 HindIII	0.45	0.28	0.23	0.45	0.47	0.95	-	0.21	0.02	0.03	0.01	0.08	0.03	0.03	0.10	0.12	0.10	0.23	0.10	0.05	0.05	0.10
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
EG1.4 HindIII	0.11	0.08	0.07	0.06	0.07	0.20	0.24	-	0.95	0.87	0.76	0.86	0.81	0.60	0.07	0.13	0.61	0.58	0.04	0.24	0.14	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	0.28	0.24	0.20
JG2E1 PstI	0.07	0.08	0.03	0.09	0.08	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
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	-	0.83	0.97	0.76	0.56	0.52	0.47	0.70	0.32	0.31	0.25	0.22
H2.8 NcoI	0.05	0.07	0.01	0.08	0.06	0.31	0.31	0.45	0.92	0.93	1.00	0.92	-	0.74	0.65	0.13	0.18	0.60	0.59	0.10	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
J44 XbaI	0.18	0.05	0.06	0.01	0.01	0.28	0.26	0.45	0.71	0.69	0.80	0.90	0.80	0.85	-	0.33	0.40	0.65	0.64	0.32	0.24	0.23
10-1X.6 AccI	0.16	0.10	0.24	0.10	0.11	0.42	0.42	0.64	0.54	0.58	0.64	0.70	0.69	0.69	0.59	-	0.91	0.19	0.36	0.56	0.00	0.02
10-1X.6 HaeIII	0.16	0.10	0.25	0.08	0.11	0.41	0.41	0.65	0.54	0.58	0.64	0.70	0.69	0.69	1.00	-	0.18	0.43	0.62	0.02	0.02	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
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
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
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	0.24	0.21	0.21	0.27	0.26	0.13	0.21	0.18	-	0.84
J3.11 MspI	0.14	0.11	0.15	0.07	0.06	0.05	0.05	0.12	0.11	0.10	0.13	0.18	0.19	0.15	0.20	0.28	0.29	0.24	0.14	0.07	0.81	-
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	0.23	0.16	0.06	0.85	0.97

CTI Chromosomen

[File: 100327527.doc // HS21K01] Übersetzung 16. August 2001
 (EP 91902210.3 / EP 0667900) DE 69132616.9-08
 HSC Research Development Corporation



Es wurde auch eine starke allelische Assoziation unter Subgruppen von RFLPs sowohl auf den CF- als auch den N-Chromosomen nachgewiesen. Wie in Tabelle 3 gezeigt, schienen die DNA-Marker, die sich physisch nahe beieinander befinden, im allgemeinen eine starke gegenseitige Assoziation zu haben. Beispielsweise wurde eine starke (in einigen Fällen fast vollständige) allelische Assoziation zwischen den benachbarten Markern E6 und E7, zwischen pH131 und W3D1.4, zwischen den polymorphen AccI- und HaeIII-Stellen, die von 10-1X.6 nachgewiesen wurden, und zwischen EG1.4, JG2E1, E2.6(E.9), E2.8 und E4.1 nachgewiesen. Die beiden Gruppen von distalen Markern in der MET- und D7S8-Region zeigten auch einen gewissen Grad an Verbindungsungleichgewicht untereinander, aber sie zeigten nur eine geringe Assoziation mit Markern von E6 bis CE1.0, was mit der unterschiedlichen Lagen von MET und D7S8 konsistent ist. Andererseits kann das Fehlen einer Assoziation zwischen DNA-Markern, die sich physisch nahe sind, auf die Existenz von Rekombinations-Hot Spots hinweisen. Beispiele für diese möglichen Hot Spots sind die Region zwischen E7 und pH131, die Umgebung von H2.3A, zwischen J44 und die Regionen, die von den Sonden 10-1X.6 und T6/20 abgedeckt werden (siehe Fig. 14). Diese Regionen, die häufige Rekombinations-Unterbrechungspunkte enthalten, waren bei der anschließenden Analyse von extendierten Haplotypdaten für die CF-Region nützlich.

20 3.5 HAPLOTYP-ANALYSE

Extendierte Haplotypen, basierend auf 23 DNA-Markern, wurden für die CF- und N-Chromosomen der Gruppe von Familien verwendet, die zuvor für die Kopplungsanalyse eingesetzt worden waren. Unter der Annahme einer Rekombination zwischen Chromosomen verschiedener Haplotypen war es möglich, mehrere Linien der beobachteten CF-Chromosomen zu konstruieren und auch die Lage des Krankheitslocus vorherzusagen.

Um weitere Informationen zu erhalten, die für das Verständnis der Natur verschiedener CF-Mutationen hilfreich sind, wurden die F508-Deletionsdaten mit den extendierten DNA-Markerhaplotypen korreliert. Wie in Tabelle 4 gezeigt, konnten vier



Hauptgruppen von N- und CF-Haplotypen durch die RFLPs definiert werden, die sich innerhalb des vermuteten CF-Gens oder unmittelbar in dessen Nachbarschaft befanden (Regionen 6-8).



TABELLE 4 (Fortsetzung)

- (a) die Daten der extendierten Halotypen werden aus den CF-Familien erhalten, die in früheren Kopplungsstudien eingesetzt worden (siehe Fußnote (a) von Tabelle 3) mit zusätzlichen Gruppen von CF-PS-Familien, die im Anschluß daran gebildet worden waren (Kerem et al., *Am. J. Genet.* 44:827 (1989)). Die Daten sind in Gruppen (Regionen) angegeben, um Platz zu sparen. Die Zuordnung der Regionen erfolgte in erster Linie gemäß paarweisen Assoziationsdaten, die in Tabelle 4 gezeigt sind, wobei die Regionen 6 – 8 den vermuteten CF-Locus umfassen (den F508-Deletion befindet sich zwischen den Regionen 6 und 7). Ein Gedankenstrich (–) ist an der Region gezeigt, wo der Haplotyp aufgrund von unvollständigen Daten oder der Unmöglichkeit einer eindeutigen Zuordnung nicht bestimmt wurde. Es sind auch alternative Haplotyp-Zuordnungen angegeben, wenn die Daten unvollständig waren. Nicht-klassifiziert umfaßt diejenigen Chromosomen mit mehr als 3 unbekanntem Zuordnungen. Die Haplotypdefinitionen für jede der 9 Regionen sind:

Region	1-	metD BanI	metD TagI	metH TagI
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 TagI	E7 TagI	pH131 HinfI	W3D1.4 HindIII
A =	1	2	2	2
B =	2	1	1	1
C =	1	2	1	1
D =	2	1	2	2
E =	2	2	2	1
F =	2	2	1	1
G =	1	2	1	2
H =	1	1	2	2

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

Region 4-	EG1.4 HincII	EG1.4 BglII	JG2E1 PstI
A =	1	1	2
B =	2	2	1
C =	2	2	2
D =	1	1	1
E =	1	2	1

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.6 AccI	10-1X.6 HaeIII
A =	1	2	1
B =	2	1	2
C =	1	1	2
D =	1	2	2
E =	2	2	2
F =	2	2	1

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

Region 8-	H1.3 NcoI	Ce1.0 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) Zahl der Chromosomen, die in jeder Gruppe indexiert wurden:

CF-PI (F) = CF-Chromosomen von CF-PI-Patienten mit der F508-Deletion;

5 CF-PS (F) = CF-Chromosomen von CF-PS-Patienten mit der F508-Deletion;



CF-PI = Andere CF-Chromosomen von CF-PI-Patienten;
CF-PS = Andere CF-Chromosomen von CF-PS-Patienten;
N = Normale Chromosomen, erhalten von Trägereltern

5 Es war offensichtlich, daß die meisten Rekombinationen zwischen Haplotypen
zwischen den Regionen 1 und 2 und zwischen den Regionen 8 und 9 auftraten, was
wiederum gut mit der langen physischen Distanz zwischen diesen Regionen überein-
stimmte. Andere, weniger häufige Unterbrechungspunkte wurden zwischen Intervallen
mit geringer Entfernung entdeckt, und sie entsprachen im allgemeinen den Hot Spots,
10 die aus den paarweisen allelischen Assoziationsstudien erhalten worden waren, wie vor-
stehend gezeigt. Es ist interessant, daß die F508-Deletion fast ausschließlich mit Gruppe
I, dem häufigsten CF-Haplotyp, assoziierte, was die Annahme unterstützt, daß diese
Deletion die Hauptmutation in CF darstellt. Wichtiger noch ist, daß, während die F508-
Deletion in 89 % (62/70) der CF-Chromosomen mit dem AA-Haplotyp (der den beiden
15 Regionen 6 und 7 entspricht), der diese Deletion flankiert, nachgewiesen wurde, sie
nicht in den 14 N-Chromosomen der gleichen Gruppe gefunden wurde ($\chi^2 = 47.3$,
 $p < 10^{-4}$). Die F508-Deletion war darum kein Sequenzpolymorphismus, der mit dem
Zentrum des Haplotyps der Gruppe I assoziiert war (siehe Tabelle 5).

20 Zusammen untermauern die Ergebnisse der Oligonucleotid-Hybridisationsstudie
und der Haplotypanalyse die Tatsache, daß es sich bei dem hier beschriebenen Genlocus
um das CF-Gen handelt, und daß die 3 bp (F508) -Deletion die häufigste Mutation in
CF darstellt.

25 3.6 INTRON/EXON-GRENZEN

Das vollständige genomische CF-Gen schließt die gesamte genetische Regula-
tionsinformation ebenso wie genetische Introninformation ein, die bei der Expression
des CF-Gens abgespleißt werden. Intronabschnitte an den Intron/Exon-Grenzen für die
30 Exons des CF-Gens sind sehr hilfreich bei der Lokalisierung von Mutationen im
CF-Gen, und sie ermöglichen die PCR-Analyse aus genomischer DNA. Genomische



DNA kann aus jedem Gewebe einschließlich der Blutleukocyten erhalten werden. Solche Introninformationen können bei der PCR-Analyse zum Zwecke des CF-Screenings verwendet werden, was in einem späteren Abschnitt ausführlicher beschrieben werden wird. Wie in Fig. 1 unter den Überschriften „Exon 1 bis Exon 24“
5 aufgeführt, gibt es Abschnitte der Bindungsintrons, insbesondere diejenigen, die die Exons flankieren, die für die PCR-Amplifizierung der Exons von grundlegender Bedeutung sind.

Eine weitere Interpretationshilfe für die Informationen in Fig. 18 ist in Fig. 21
10 angegeben. Es werden genomische DNA-Klone angegeben, die die Kodierungsregion des CFTR-Gens enthalten. Wie aus Fig. 21 deutlich wird, bestehen beträchtliche Lücken zwischen den Klonen der Exons, die auf Lücken in den Intronabschnitten zwischen den Exons der Fig. 18 hinweisen. Diese Lücken in den Intronabschnitten sind mit „...“ bezeichnet. In Fig. 21 wurden die Klone mittels verschiedener Endonucleasen kartiert
15 (AccI,A; AvaI,W; AamHI,B; BgIII,G; BssHI,Y; EcoRV,V; FspI,F; HincII,C; HindIII,H; KpnI,K; NcoI,J; PstI,P; PvuII,U; SmaI,M; SacI,S; SspI,E; StyI,T; XbaI,X; XhoI,O). In Fig. 21 sind die Exons durch Kästchen dargestellt. Die ungefüllten Kästchen zeigen nicht-kodierende Abschnitte der Exons an, wogegen die gefüllten Kästchen kodierende Abschnitte anzeigen. Die vermutlichen Positionen der Exons innerhalb der genomischen
20 DNA sind ebenfalls anhand ihrer relativen Positionen angegeben. Die Pfeile über den Kästchen markieren die Lage der Oligonucleotide, die als Sequenzierungsprimer bei der PCR-Amplifikation der genomischen DNA verwendet wurden. Die Zahlen, die unter der Restriktionskarte angegeben sind, geben die Größe des Restriktionsfragments in kb wieder.

25

Bei der Sequenzierung der Intronabschnitte wurde festgestellt, daß es mindestens
27 Exons anstelle der zuvor in den vorstehend genannten mit-anhängigen Anmeldungen angegebenen 24 Exons gibt. Die Exons 6, 14 und 17 finden sich, wie bereits berichtet wurde, in Segmenten und werden nun als Exon 6a, 6b, Exon 14a, 14b und Exon 17a,
30 17b bezeichnet.



Die Intronabschnitte, die bei der PCR-Amplifikation verwendet wurden, sind in der folgenden Tabelle 5 identifiziert und in Fig. 18 unterstrichen. Die von den Pfeilen bezeichneten Abschnitte werden selektiert, aber selbstverständlich sind auch andere Abschnitte der Intronsequenzen im PCR-Amplifikationsverfahren nützlich. Beispielsweise wird für Exon 10 die relevante genetische Information, die bei der PCR bevorzugt ist, mit Bezug auf die 5'- und 3'-Enden der Sequenz gekennzeichnet. Die Intronsektion ist mit einem „i“ bezeichnet. Daher sind in Tabelle 5 die bevorzugten Abschnitte für Exon 2 mit 2i-5 und 2i-3 und ähnlich für die Exon 3 bis 24 bezeichnet. Bei Exon 1 enthält der ausgewählte Abschnitt die Sequenz GGA...AAA für B115-B und ACA...GTG für 10D. Bei Exon 13 werden die Abschnitte durch zwei Gruppen identifiziert: 13i-5 und C1-1m und Y13B-5 und 13i-3A. (Dieses Exon (13) ist groß und es ist am praktischsten, es aus zwei Abschnitten zusammensetzen). C1-1M und X13B-5 stammen aus Exonsequenzen. Die speziellen Bedingungen für die PCR-Amplifikation der einzelnen Exons sind in der folgenden Tabelle 6 aufgeführt und ausführlicher im folgenden mit Bezug auf die in R.K. Saiki et al., *Science* 230:1350 81985) erklärten Vorgehensweisen erörtert.

Diese Oligonucleotide, wie aus der Intronsequenz erhalten, unterstützen die Amplifikation des jeweiligen Exons durch PCR, wobei sie für die Analyse bezüglich DNA-Sequenzveränderungen sorgen, die Mutationen auf dem CF-Gen entsprechen. Die Mutationen können entweder durch direkte Sequenzbestimmung der PCR-Produkte oder durch Sequenzieren der Produkte, die in Plasmidvektoren kloniert wurden, gezeigt werden. Das amplifizierte Exon kann auch mittels Gelelektrophorese auf eine Weise, die näher zu beschreiben ist, analysiert werden. Es hat sich gezeigt, daß die Abschnitte des Introns für das jeweilige Exon genügend lang sind, um sich besonders gut für die PCR-Technik zu eignen, um eine Amplifikation des relevanten Exons zu liefern.

68 64

TABELLE 5

Oligonucleotide, verwendet zur Amplifikation von CF-Gen-Exons durch PCR

Exon	PCR-Primer; 5'→3'	Amplifiziertes Product (bp)
1	GGAGTTCACCTACCTAAA (B115-B) ACACGCCCTCCTCTTTTCGTG (10D)	933
2	CCAAATCTGTATGGAGACCA (2i-5) TATGTTGCCAGGCTGGTAT (2i-3)	378
3	CTGGGTTAATCTCCTTGG (3i-5) ATTCACCAGATTTCGTAGTC (3i-3)	309
4	TCACATATGGTATGACCCTC (4i-5) TTGTACCAGCTCACTACCTA (4i-3)	438
5	ATTTCTGCCTAGATGCTGGG (5i-5) AACTCCGCCTTCCAGTTGT (5i-3)	395
6a	TTAGTGTGCTCAGAACCACG (6Ai-5) CTATGCATAGAGCAGTCCTG (6Ai-3)	385
6b	TGGAATGAGTCTGTACAGCG (6Ci-5) GAGGTGGAAGTCTACCATGA (6Ci-3)	417
7	AGACCATGCTCAGATCTTCCAT (7i-5) GCAAAGTTCATTAGAACTGATC (7i-3)	410
8	TGAATCCTAGTGCTTGCAA (8i-5) TCGCCATTAGGATGAAATCC (8i-3)	359
9	TAATGGATCATGGGCCATGT (9i-5) ACAGTGTGAATGTGGTGCA (9i-3)	560
10	GCAGAGTACCTGAAACAGGA (10i-5) CATTCACAGTAGCTTACCCA (10i-3)	491
11	CAACTGTGGTTAAAGCAATAGTGT (11i-5) GCACAGATTCTGAGTAACCATAAT (11i-3)	425
12	GTGAATCGATGTGGTGACCA (12i-5) CTGGTTTAGCATGAGGCGGT (12i-3)	426
13 (a)	TGCTAAAATACGAGACATATTGCA (13i-5) ATCTGGTACTAAGGACAG (C1-1M)	528
(b)	TCAATCCAATCAACTCTATACGAA (X13B-5) TACACCTTATCCTAATCCTATGAT (13i-3A)	497
14a	AAAAGGTATGCCACTGTAAAG (14Ai-5) GTATACATCCCCAACTATCT (14Ai-3)	511
14b	GAACACCTAGTACAGCTGCT (14Bi-5) AACTCCTGGGCTCAAGTGAT (14Bi-3)	449
15	GTGCATGCTCTTCTAATGCA (15i-5) AAGGCACATGCCTCTGTGCA (15i-3)	485
16	CAGAGAAAATTGGTGTACT (16i-5) ATCTAAATGTGGGATTGCCT (16i-3)	570
17a	CAATGTGCACATGTACCCTA (17Ai-5) TGTACACCAACTGTGGTAAG (17Ai-3)	579
17b	TTCAAAGAATGGCACCAGTGT (17Bi-5) ATAACCTATAGAATGCAGCA (17Bi-3)	463
18	GTAGATGCTGTGATGAACTG (18i-5) AGTGGCTATCTATGAGAAGG (18i-3)	451
19	GCCCGACAAATAACCAAGTGA (19i-5) GCTAACACATTGCTCAGGCT (19i-3)	454
20	GGTCAGGATTGAAAGTGTGCA (20i-5) CTATGAGAAAACCTGCACTGGA (20i-3)	473
21	AATGTTTACAAGGGACTCCA (21i-5) CAAAAGTACCTGTGTGCTCCA (21i-3)	477
22	AAACGCTGAGCCTCACAAGA (22i-5) TGTCACCATGAAGCAGGCAT (22i-3)	562
23	AGCTGATTGTGCGTAACGCT (23i-5) TAAAGCTGGATGGCTGTATG (23i-3)	400
24	GGACACAGCAGTTAAATGTG (24i-5) ACTATTGCCAGGAAGCCATT (24i-3)	569



TABELLE 6

Exon	Puffer ^a	Anfängliche			Endgültige	
		Denaturierung	Denaturierung	Erhitzung	Extension	Extension
		Dauer/Temp.	Dauer/Temp.	Dauer/Temp	Dauer/Temp.	Dauer/Temp.
3-5, 6a, 6b, 1 bis 10,12, 14a, 16, 17b, 18-24	A (1,5)	6 min/94°C	30 s/94°C	30 s/55°C	1 min/72°C	7 min/72°C
1	B	6 min/94°C	30 s/94°C	30 s/55°C	2,5 min/72°C	7 min/72°C
2,11	B	6 min/94°C	30 s/94°C	30 s/55°C	1 min/72°C	7 min/72°C
13a	A(1,75)	6 min/94°C	30 s/94°C	30 s/55°C	2,5 min/72°C	7 min/72°C
13b	A(1,75)	6 min/94°C	30 s/94°C	30 s/55°C	2,5 min/72°C	7 min/72°C
14b	B	6 min/94°C	30 s/94°C	30 s/55°C	1 min/72°C	7 min/72°C
17a	A(1,5)	6 min/94°C	30 s/94°C	30 s/55°C	1 min/72°C	7 min/72°C

(a) Puffer A(1.5) : * Puffer mit 1,5 mM MgCl₂

5 Puffer A(1.75): * Puffer mit 1,75 mM MgCl₂

Puffer B : 67 mM Tris-HCl pH 8,8, 6,7 mM MgCl₂, 16,6 mM (NH₄)₂SO₄, 0,67
 µM EDTA, 10 mM B-Mercaptoethanol, 170 µg/ml BSA, 10 %
 DMSO, 1,5 mM von jedem dNTP

* der Puffer A enthält: 10 mM Tris pH 8,3 (bei 25°C)

10 0,001 % (Gew./Gew.) Gelatine

0,2 mM dNTPs

dNTPs = Desocynucleotidtriphosphate

2.7 CF-MUTATIONEN – ΔI506 oder ΔI507

15

Die Assoziation der F508-Deletion mit 1 üblichen und 1 seltenen CF-Haplotyp lieferte weitere Einsichten in die Zahl der Mutationsgeschehnisse, die zur gegenwärtigen Patientenpopulation beitragen. Aufgrund von umfangreichen Haplotypdaten ist es wahrscheinlich, daß das ursprüngliche Chromosom, in dem die F508-Deletion auftrat, den Haplotyp –AAAAAAA– trägt (Gruppe Ia), wie in Tabelle 4 definiert. Die
 20 anderen CF-Chromosomen der Gruppe I, die die Deletion tragen, sind wahrscheinlich



Rekombinationsprodukte, die vom ursprünglichen Chromosom abgeleitet wurden. Wenn man davon ausgeht, daß die CF-Chromosomen in jeder Haplotypgruppe den gleichen Ursprung haben, würden nur 3 – 4 zusätzliche Mutationsgeschehnisse voraus-
5 gesagt (siehe Tabelle 4). Da jedoch viele der CF-Chromosomen in der gleichen Gruppe sich deutlich voneinander unterscheiden, ist eine weitere Unterteilung innerhalb jeder Gruppe möglich. Infolgedessen kann von einer höheren Zahl von unabhängigen Mutationsgeschehnissen ausgegangen werden, und die Daten legen nahe, daß mindestens 7 zusätzliche vermutete Mutationen ebenfalls zum CF-PI-Phänotyp beitragen (siehe Tabelle 3). Die Mutationen, die zur CF-PS-Untergruppe führen, sind wahrschein-
10 lich heterogener.

Die 7 zusätzlichen CF-PI-Mutationen werden von den Haplotypen: –CAAAAAA– (Gruppe Ib), –CABCAAD– (Gruppe Ic), – –BBBAC– (Gruppe Ia), –CABBBAB– (Gruppe Va) dargestellt. Obwohl der Molekulardefekt in jeder dieser Mutationen noch
15 definiert werden muß, ist es klar, daß keine dieser Mutationen die Region, die den Oligonucleotid-Bindungsstellen entspricht, die im PCR/Hybridisierungs-Versuch verwendet werden, ernsthaft beeinflußt.

Es wurde jedoch gefunden, daß ein CF-Chromosom, das an die $\Delta F508$ -ASO-
20 Sonde hybridisierte, mit einem anderen Haplotyp (Gruppe IIIa) assoziierte. Es schien, daß $\Delta F508$ in beiden Haplotypenaufreten sollte, aber mit der Entdeckung von $\Delta I507$ wurde nachgewiesen, daß dies nicht der Fall ist. Stattdessen befindet sich $\Delta F508$ in Gruppe Ia, wogegen $\Delta I507$ sich in Gruppe IIIa findet. Keines der anderen CF- oder normalen Chromosomen dieser Haplotypgruppe (IIIa) hat eine Hybridisation an den
25 Mutanten ($\Delta F508$) ASO gezeigt [B. Kerem et al., *Science* 245:1073 (1989)]. Angesichts dessen, daß Haplotypen der Gruppen Ia und IIIa sich deutlich voneinander unterscheiden, müssen sich die Mutationen, die von diesen beiden CF-Chromosomengruppen beherbergt werden, unabhängig voneinander entwickelt haben. Um die molekulare Natur der Mutation in diesen CF-Chromosomen der Gruppe III zu untersuchen, be-
30 schrieben wir die uns interessierende Region durch Amplifikation der genomischen



DNA aus einem Individuum, das das Chromosom IIIa trug, durch Polymerase-Chain-reaction (PCR).

Diese Polymerase-Chainreactions (PCR) wurden gemäß dem Verfahren von
5 R.K. Saiki et al., *Science* 230:1350 (1985) durchgeführt. Ein spezifisches DNA-Segment aus 491 bp einschließlich des Exons 10 des CF-Gens wurden unter Verwendung der Oligonucleotidprimer 10i-5 (5'-GCAGAGTACCTGAAACAGGA-3') und 10i-5 (5'-GCAGAGTACCTGAAACAGGA-3'), die sich in den 5'- bzw. 3' flankierenden Regionen befanden, wie in Fig. 18 gezeigt und in Tabelle 5 im einzelnen aufgeführt.
10 Beide Oligonucleotide wurden vom HSC DNA Biotechnology Service Center (Toronto) erworben. Ungefähr 500 ng der genomischen DNA aus kultivierten lymphoblastoiden Zelllinien der Eltern und des CF-Kindes der Familie 5 wurden in jeder Reaktion verwendet. Die DNA-Proben wurden in einem Perkin-Elmer/Cetus DNA Thermal Cycler bei 94°C 30 s lang denaturiert, die Primer wurden bei 55°C 30 s lang erhitzt und bei
15 72°C 50 s lang (mit 0,5 Einheiten Taq-Polymerase, Perkin-Elmer/Cetus, Norwalk, CT) mit 30 Durchgängen und einer abschließenden Extensionsperiode von 7 min extendiert. Die Reaktionsbedingungen für die PCR-Amplifikation anderer Exons sind in Tabelle 6 angegeben.

20 Es wurde eine Hybridisationsanalyse der PCR-Produkte aus drei Individuen der Familie 5 aus Gruppe IIIa durchgeführt. Fig. 19a: die Trägermutter und der Trägervater sind durch einen halbgefüllten Kreis bzw. ein halbgefülltes Viereck angegeben und der erkrankte Sohn ist mit einem gefüllten Viereck angegeben. Die Bedingungen für die Hybridisation und das Waschen wurden bereits beschrieben (Kerem et al., s.o.). Im
25 PCR-Produkt des Vaters gibt es ein ziemlich schwaches Signal mit der Mutanten (oligo $\Delta F508$) -Sonde. In Fig. 19b sind die DNA-Sequenzanalyse des Klons 5-3-15 und die PCR-Produkte aus dem erkrankten Sohn und dem Trägervater gezeigt. Der Pfeil in der mittleren Tafel zeigt die Anwesenheit von sowohl A- als auch T-Nucleotidresten an der gleichen Position an; der Pfeil in der rechten Tafel zeigt die Divergenzpunkte zwischen
30 der normalen und der $\Delta I507$ -Sequenz an. Die gezeigten Sequenzleitern wurden aus den umgekehrten Komplementen abgeleitet, wie später beschrieben wird. Fig. 19c zeigt die



DNA-Sequenzen und ihre entsprechenden Aminosäuresequenzen der normalen, $\Delta I507$ und $\Delta F508$ -Allele, die die Mutationsstellen umfassen, wie gezeigt. Fig. 19a: die PCR-amplifizierte DNA aus dem Trägervater, der das CF-Chromosom der Gruppe IIIa auf den erkrankten Sohn übertragen hatte, hybridisierte weniger effizient mit $\Delta F508$ ASO als dasjenige von der Mutter, die das CF-Chromosom der Gruppe Ia trug. Der Unterschied wurde deutlich, als die Hybridisationssignale mit denen mit der normalen ASO-Sonde verglichen wurden. Dieses Ergebnis zeigte daher an, daß die Mutation, die vom CF-Chromosom der Gruppe IIIa getragen wurde, möglicherweise nicht mit $\Delta F508$ identisch ist.

10

Um die Nucleotidsequenz zu definieren, die dem Mutantenallel auf diesem Chromosom entspricht, wurde das PCR-amplifizierte Produkt der DNA des Vaters aus einem Polyacrylamid-Elektrophoresegel excisiert und in einen Sequenziervektor kloniert.

15

Die allgemeine Vorgehensweise für die DNA-Isolierung und Reinigung für die Klonierung in einen Sequenziervektor sind in J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2. Auflage (Cold Spring Harbor Press, N.Y. 1989) beschrieben. Die beiden Homoduplexe, die durch die PCR-Amplifikation der elterlichen DNA erzeugt worden waren, wurden von einem 5%-igen nicht-denaturierenden Polyacrylamidgel (30:1 Acrylamid:Bisacrylamid) gereinigt. Die geeigneten Banden wurden durch Anfärben mit Ethidiumbromid sichtbar gemacht, excisiert und in TE (10 mM Tris-HCl; 1 mM EDTA; pH 7,5) 2 bis 12 Stunden lang bei Raumtemperatur eluiert. Die DNA-Lösung wurde nacheinander mit Tris-equilibriertem Phenol, Phenol/ $CHCl_3$ und $CHCl_3$ behandelt. Die DNA-Proben wurden durch Ausfällen in Ethanol und Resuspension in TE aufkonzentriert, mit T4-Polynucleotidkinase in Anwesenheit von ATP inkubiert und in diphosphoryliertem, stumpfendigem Bluescript KSTM-Vektor (Stratagene, San Diego, CA) ligiert. Klone, die amplifiziertes Produkt enthielten, das aus dem normalen elterlichen Chromosom erzeugt worden war, wurden durch Hybridisation mit dem Oligonucleotid N wie von Kerem et al., s.o., beschrieben identifiziert.

30



Klone, die die Mutantensequenz enthielten, wurden anhand ihrer Unfähigkeit, den normalen ASO zu hybridisieren, identifiziert (Kerem et al., s.o.). Ein Klon, 5-3-15, wurde isoliert und seine DNA-Sequenz wurde bestimmt. Das allgemeine Protokoll zum Sequenzieren geklonter DNA ist im wesentlichen wie bereits beschrieben [J.R. Riordan et al., *Science* 245:1066 (f1989)], wobei ein U.S. Biochemicals Sequenase™-Kit verwendet wird. Um die Sequenz zu verifizieren und um jeglichen Irrtum auszuschließen, der durch Polymerase während der PCR eingebracht werden könnte, wurden die DNA-Sequenzen für die PCR-Produkte vom Vater und einem der erkrankten Kinder ebenfalls direkt ohne Klonieren bestimmt.

Dieses Verfahren wurde durch Denaturieren von 2 pMol Gel-gereinigtem doppelsträngigem Produkt in 0,2 M NaOH/0,2 mM EDTA (5 min bei Raumtemperatur) durchgeführt, durch Zugabe von 0,1 Volumina 2 M Ammoniumacetat (pH 5,4) neutralisiert und mit 2,5 Volumina Ethanol 10 min lang bei -70°C ausgefällt. Nach dem Waschen mit 70 % Ethanol wurde das DNA-Pellet getrocknet und in einem Sequenzierreaktionspuffer, der 4 pMol des Oligonucleotidprimers 10i-3 von Fig. 18, Dithiothreitol (8,3 mM) und [α -35S]-dATP (0,8 μ M, 1000 Ci/mMol) enthielt, erneut gelöst. Die Mischung wurde bei 37°C 20 min lang inkubiert, wonach 2 μ l Markierungsmischung, wie im Sequenase™ Kit enthalten, und dann 2 Einheiten Sequenaseenzym zugegeben wurden. Aliquote der Reaktionsmischung (3,5 μ l) wurden ohne Verzögerung in Röhren übertragen, von denen jede 2,5 μ l ddGTP-, ddATP-, ddTTP- und ddCTP-Lösungen enthielt (U.S. Biochemicals Sequenase Kit) und die Reaktionen wurden durch Zugabe der Stopplösung gestoppt.

25

Die DNA-Sequenz für dieses Mutantenallel ist in Fig. 19b gezeigt. Die Daten, die von der klonierten DNA und dem direkten Sequenzieren des PCR-Produkts des erkrankten Kindes und des Vaters abgeleitet wurden, waren alle mit einer 3 bp-Deletion konsistent, wenn man sie mit der normalen Sequenz verglich (Fig. 19c). Die Deletion dieses 3 bp (ATC) an der I506- oder I507-Position resultiert in dem Verlust eines Iso-leucinrests aus der vermuteten CFTR, innerhalb der gleichen ATP-Bindungsdomäne, in

30



sich $\Delta F508$ befindet, aber es ist nicht klar, ob diese deletierte Aminosäure der Position 506 oder 507 entspricht. Da die Positionen 506 und 507 Wiederholungen sind, ist es gegenwärtig nicht möglich, zu bestimmen, an welcher Position die 3 bp-Deletion auftritt. Um spätere Erörterungen zu vereinfachen, bezeichnen wir diese Deletion jedoch als
5 $\Delta I507$.

Die Tatsache, daß die $\Delta I507$ - und $\Delta F508$ -Mutationen in der gleichen Region der vermuteten ATP-Bindungsdomäne von CFTR auftreten, ist überraschend. Obwohl nicht die gesamte Sequenz des $\Delta I507$ -Allel untersucht wurde, wie dies bei $\Delta F508$ der Fall war, deutet die strategische Lage der Deletion darauf hin, daß dies die für dieses Allel verantwortliche Mutation ist. Dieses Argument wird außerdem durch die Beobachtung unterstützt, daß diese Veränderung nicht in einem der normalen Chromosomen nachgewiesen wurde, die bisher untersucht wurden. (Kerem et al., s.o.). Die Identifizierung einer zweiten einzelnen Aminosäuredeletion in der ATP-Bindungsdomäne von CFTR
15 liefert ebenfalls Informationen über die Struktur und Funktion dieses Proteins. Da die Deletion entweder des Phenylalaninrests an der Position $\Delta I508$ oder des Isoleucins an der Position $\Delta I507$ ausreicht, um die Funktion von CFTR zu beeinflussen, so daß die CF-Krankheit verursacht wird, nimmt man an, daß diese Reste an der Faltung des Proteins, aber nicht direkt an der Bindung des ATP beteiligt sind. Das heißt, die Länge
20 des Peptids ist möglicherweise wichtiger als der jeweilige Aminosäurerest in dieser Region. Als Unterstützung für diese Hypothese wurde gefunden, daß, ohne sichtbaren Funktionsverlust von CFTR, der Phenylalaninrest durch ein Serin, und Isoleucin an der Position 506 durch Valin ersetzt werden können.

Als die Nucleotidsequenz von $\Delta I507$ mit der von $\Delta F508$ in der ASO-Hybridisierungsregion verglichen wurde, wurde festgestellt, daß der Unterschied zwischen den beiden Allelen nur in einem Wechsel $A \rightarrow T$ bestand (Fig. 19c). Dieser kleine Unterschied erklärte somit die Kreuzungshybridisation des $\Delta R508$ -ASO zu $\Delta I507$. Diese Ergebnisse veranschaulichten daher die Bedeutung einer sorgfältigen Prüfung beider
30 Elternchromosome bei der Durchführung der auf ASO basierenden genetischen Diagnose. Es wurde festgestellt, daß die $\Delta F508$ - und $\Delta I507$ -Mutationen durch Erhöhen der



Stringenz der Oligonucleotidhybridisierungs-Bedingungen oder durch Nachweis der
einzigartigen Mobilität der Heteroduplexe, die zwischen jeder dieser Sequenzen und der
normalen DNA auf einem Polyacrylamidgel gebildet wurden, unterschieden werden
können. Die Stringenz der Hybridisierung kann durch Verwendung einer Wasch-
5 temperatur von 45°C anstelle der vorherigen 39°C in Anwesenheit von 2XSSC
(1XSSC = 150 mM NaCl und 15 mM Na-Citrat) erhöht werden.

Die Identifizierung der $\Delta I507$ - und $\Delta F508$ -Allele durch Polyacrylamidgel-Elektro-
phorese ist in Fig. 20 gezeigt. Die PCR-Produkte wurden aus den drei Familien-
10 mitgliedern erzeugt und auf einem 5 % Polyacrylamidgel, wie vorstehend beschrieben,
separiert. Eine DNA-Probe von einem bekannten heterocygoten $\Delta R508$ -Träger ist aus
Vergleichsgründen eingeschlossen. Fig. 20: Das Bandenmuster der PCR-amplifizierten
genomischen DNA des Vaters, der der Träger von $\Delta I507$ ist, läßt sich deutlich von dem
der Mutter unterscheiden, die zu dem Trägertyp mit der $\Delta F508$ -Mutation gehört. An
15 diesem Gelelektrophoresetest nahmen eigentlich drei Individuen teil (der Trägervater
und die beiden erkrankten Söhne in Familie 5), die die $\Delta I507$ -Deletion trugen. Da sie
alle zur gleichen Familie gehören, repräsentieren sie in unser Populationsanalyse nur ein
einziges CF-Chromosom [Kerem et al., s.o.]. Die beiden Patienten, die auch die $\Delta F508$ -
Mutation von ihrer Mutter geerbt hatten, zeigte typische CF-Symptome mit Bauch-
20 speicheldrüsen-Unterfunktion. Der Vater dieser Familie war das einzige Elternteil, das
diese $\Delta I507$ -Mutation trug; keiner der anderen Eltern zeigte in der retrospektiven Studie
ein verringertes Hybriditäts-Intensitätssignal mit der $\Delta F508$ -Oligonucleotid-Mutanten-
sonde oder ein besonderes Heteroduplexmuster für das PCR-Produkt (wie vorstehend
definiert). Darüber hinaus wurden zwei Vertreter der CF-Chromosomen-Gruppe IIIb
25 und einer der Gruppe IIIc aus unserer Sammlung [Kerem et al., s.o.] sequenziert, aber es
zeigte sich, daß keiner $\Delta I507$ enthielt. Da die Elektrophoresetechnik die Notwendigkeit
für Sondenmarkierung und Hybridisation überflüssig macht, kann sie sich als die
Methode der Wahl zum Nachweisen von Trägern in einem großen Populationsmaßstab
erweisen [J.M. Rommens et al., *Am. J. Hum. Genet.* 46:395-396 81990)].

30



Die gegenwärtigen Daten zeigen auch an, daß eine strikte Korrelation zwischen dem Haplotyp des DNA-Markers und der Mutation in CF besteht. Die $\Delta F508$ -Deletion ist die häufigste CF-Mutation, die vor dem Hintergrund eines Chromosoms der Gruppe Ia auftrat [Kerem et al. s.o.]. Die $\Delta I507$ -Mutation ist jedoch in der CF-Population
5 selten; das eine CF-Chromosom der Gruppe IIIa, das diese Deletion trug, ist das einzige Beispiel in unserer untersuchten Population (1/219). Da der Haplotyp der Gruppe III unter den normalen Chromosomen relativ weit verbreitet ist (17/198), entstand die $\Delta I507$ -Deletion vermutlich erst vor kurzem. Weitere Studien mit größeren Populationen mit unterschiedlichem geographischem und ethnischem Hintergrund sollten weitere
10 Erkenntnisse für das Verständnis der Ursprünge dieser Mutationen bringen.

3.8 WEITERE CF-MUTATIONEN

Gemäß den vorstehend genannten Verfahren wurden weitere Mutationen im CF-
15 Gen identifiziert. Die folgende kurze Beschreibung jeder identifizierten Mutation beruht auf kürzlich beschriebenen Verfahren zur Lokalisierung der Mutation unter Verwendung von PCR-Verfahren. Die in diesem Abkürzungen verwendete Numerierung bezeichnet entweder die DNA-Sequenz- oder die Aminosäuresequenz-Position der Mutation, abhängig vom Mutationstyp. Beispielsweise werden Spleißmutationen und
20 Rasterverschiebungsmutationen mittels der DNA-Sequenzposition definiert. Die meisten anderen Mutationen leiten ihre Nomenklatur von der Aminosäurerest-Position ab. Die Beschreibung jeder Mutation erklärt jedenfalls die Nomenklatur.

Beispielsweise führen die Mutationen G542X, Q493X, 3659 del C, 556 del A zu
25 verkürzten Polypeptiden, die sich deutlich von den Deletionen oder Veränderungen einzelner Aminosäuren unterscheiden. G542X und Q493X beinhalten ein Polypeptid, einschließlich der ersten 541 bzw. 493 Aminosäurereste des normalen 1480-Aminosäurepolypeptids. 3659 del C und 556 del A beinhalten ebenfalls gekürzte Versionen und enthalten zusätzliche Aminosäurereste. Es wird vorausgesagt, daß die Mutationen
30 711+1G \rightarrow T und 1717 - 1G \rightarrow A zu Polypeptiden führen, die bisher nicht exakt definiert werden konnten. Sie führen wahrscheinlich nicht zu gekürzten Polypeptiden,



sondern könnten zusätzliche Aminosäuren enthalten. Die DNA-Sequenz, die für diese Mutantenpolypeptide kodiert, wird vermutlich eine Intronsequenz aus dem normalen Gen oder mögliche deletierte Exons enthalten.

5 3.8.0 MUTATIONEN IM EXON 1

Bei der 129G → C-Mutation kommt es zum Austausch eines einzigen Basenpaars von G zu C am Nucleotid 129 der cDNA-Sequenz de Fig. 1. Das PCR-Produkt zum Amplifizieren genomischer DNA, die diese Mutation enthält, wird von den B115-B und
10 10D-Primern abgeleitet, wie in Tabelle 5 dargelegt ist. Die genomische DNA wird anhand der Bedingungen amplifiziert, die in Tabelle 6 aufgelistet sind.

3.8.1 MUTATIONEN IN EXON 3

15 Die G85E-Mutation in Exod 3 beinhaltet einen Übergang von G zu A an der Nucleotidposition 386. Sie wird in der Familie Nr. 26 nachgewiesen, einer franko-kanadischen Familie, die als PI klassifiziert wurde. Dieser vorausgesagte Aminosäurewechsel von Gly zu Glu ist mit einem Haplotyp der Gruppe Iib assoziiert. Die Mutation zerstört eine Hinfl-Stelle. Das PCR-Produkt, das aus den 3i-5- und 3i-3-Primern abgeleitet wurde, nach den Bedingungen der Tabelle 6, wird durch dieses Enzym in 3 Frag-
20 mente aufgespalten, 172, 105 bzw. 32 bp, für die normale Sequenz; ein Fragment von 277 bp würde bei der Mutantensequenz vorliegen. Wir analysierten 54 CF-Chromosomen, 8 aus Gruppe II, und 50 normale Chromosomen, 44 aus Gruppe II und fanden kein anderes Beispiel für G85E.

25

3.8.2 MUTATIONEN IN EXON 4

556 del A ist eine Frameshiftmutation in Exon 4 in einem einzelnen Chromosom (Familie Nr. 17 aus Toronto, GM1076). Es existiert eine Deletion von A an der
30 Nucleotidposition 556. Diese Mutation ist mit dem Haplotyp der Gruppe IIIb assoziiert und findet sich in keinem der anderen 31 CF-Chromosomen (9 aus IIIb) und 30 N-



Chromosomen (f16 aus IIIb). Die Mutation erzeugte eine BgII Spaltungsstelle für 1 Enzym. Die PCR-Primer sind 4i-5 und 4i-3 (siehe Tabelle 5) wo das Enzym das Mutanten-PCR-Produkt (437 bp) in 2 Fragmente mit einer Größe von 287 und 150 bp zerschneidet.

5

Die I148T-Mutation in Exon 4 beinhaltet einen Basenpaar-Übergang von T nach C an der Nucleotidposition 575. Dies resultiert in einem Wechsel von Ile zu Thr an der Aminosäureposition 148 der Fig. 1. Das PCR-Produkt, das verwendet wurde, um genomische DNA, die diese Mutation enthält, zu amplifizieren, verwendet die Primer 4i-5 und 4i-3, wie in Tabelle 5 dargestellt ist. Die Reaktionsbedingungen zum Amplifizieren der genomischen DNA sind in Tabelle 6 dargestellt.

3.8.3 MUTATIONEN IN EXON 5

15 In der Mutation G178R hat die Fehlsinnmutation von Gly zu Arg in Exon 5 ihre Ursache in einer Veränderung an der Nucleotidposition 644. Die Mutation findet sich auf dem CF-Chromosom der Mutter der Familie Nr. 50; bei der anderen Mutation in dieser Familie handelt es sich um $\Delta F508$. Die Primer 5i-5 und 5i-3 wurden zum Amplifizieren der genomischen DNA verwendet, wie in den Tabellen 5 und 6 dargestellt ist.

20

3.8.4 MUTATIONEN IN EXON 9

Bei einer Mutation im Exon 9 handelt es sich um eine Veränderung von Alanin (GCG) zu Glutaminsäure (GAG) an der Aminosäureposition 455 (A455 \rightarrow E). Zwei der 38 untersuchten, nicht- $\Delta F508$ CF-Chromosomen tragen diese Mutation; beide stammen von Patienten frankokanadischer Abstammung, die in unserer Arbeit als Familien Nr. 27 und Nr. 53 identifiziert wurden, und gehören zur Haplotypgruppe Ib. Die Mutation ist durch Allel-spezifische Oligonucleotid (ASO) -Hybridisation mit PCR-amplifizierter genomischer DNA-Sequenz nachweisbar. Bei den PCR-Primern handelt es sich um 9i-5 (5'-TAATGGATCATGGGCCATGT-3') zur Amplifizierung der genomischen

30



DNA unter den Bedingungen der Tabelle 6. Bei den ASOs handelt es sich um
5'-GGTTGTTGGCGGTTGCT-3' für das normale Allel und
5'-GTTGTTGGAGGTTGCT-3' für den Mutanten. Die Oligonucleotid-Hybridisation
wird wie in Kerem et al. (1989), s.o., beschrieben, bei 37°C durchgeführt, und die
5 Waschungen werden zweimal mit 5XSSC jeweils 10 min lang bei Raumtemperatur
durchgeführt, gefolgt von zweimal mit 2 X SSC für jeweils 30 min bei 52°C. Obwohl
das Alanin an Position 455 (Ala455) nicht in allen ATP-Bindungsfalten in allen Spezies
vorhanden ist, ist es in allen bekannten Mitgliedern der P-Glycoproteinfamilie, dem
CFTR ähnlichsten Protein, vorhanden. Außerdem nimmt man an, daß A455 → E eher
10 eine Mutation als ein Sequenzpolymorphismus ist, da sich diese Änderung nicht in 16
nicht-ΔF508 CF-Chromosomen und drei normalen Chromosomen, die den gleichen
Haplotyp der Gruppe I tragen, findet.

3.8.5 MUTATIONEN IM EXON 10

15 Bei der Q493X-Mutation wird bei der Familie Nr. 9 aus Toronto Gln493 (CAG) in
einen Stopcodon (TAG) umgewandelt (Nucleotidposition 1609 C → T). Die Mutation
tritt auf einem CF-Chromosom mit dem Haplotyp IIIb auf; sie findet sich nicht in 28
normalen Chromosomen (von denen 15 zu 11b gehören) noch in 33 anderen Chromo-
20 somen (5 davon IIIb). Die Mutation kann durch Allel-spezifische PCR nachgewiesen
werden, mit 10i-5 als dem gebräuchlichen PCR-Primer,
5'-GGCATAATCCAGGAAAAGT-3' für die normale Sequenz und
5'-GGCATAATCCAGGAAAAGTA-3' für das Mutantenallel. Die PCR-Bedingung ist
6 min bei 94°, gefolgt von Durchgängen von 30 s bei 94°, 30 s bei 57° und 90 s bei 72°,
25 mit 100 ng jedes Primers und ~ 400 ng genomischer DNA. Die Primer 9i-3 und 9i-5
können für die interne PCR-Kontrolle verwendet werden, da sie die gleichen Reaktions-
bedingungen aufweisen.



3.8.6 MUTATIONEN IM EXON 11

Bei der Mutation G542X wird der Glycincodon (GGA) an der Aminosäureposition 542 in einen Stopcodon (TGA) verändert (G542 → Stop). Das einzelne Chromosom, das diese Mutation trägt, ist von ashkenasisch/jüdischem Ursprung (Familie A) und weist den Haplotyp B auf (XV2C-Allel 1; KM.19-Allel 2). Die Mutantensequenz kann durch Hybridisationsanalyse mit Allel-spezifischen Oligonucleotiden (ASOs) auf genomischer DNA, amplifiziert unter Bedingungen von Tabelle 6, durch PCR mit den 11i-5 und 11i-3-Oligonucleotid-Primern nachgewiesen werden. Bei dem normalen ASO handelt es sich um 5'-ACCTTCTCCAAGAACT-3' und bei dem Mutanten-ASO um 5'-ACCTTCTCAAAGAACT-3'. Die Oligonucleotidhybridisations-Bedingung ist wie in Kerem et al. (1989), s.o. beschrieben und die Waschbedingungen sind zweimal in 5 x SSC für 10 min, jeweils bei Raumtemperatur, gefolgt von zweimal in 2 X SSC für 30 min, jeweils bei 45°C. Die Mutation wird nicht in 52 anderen nicht-ΔF508-Chromosomen nachgewiesen, von denen 11 jüdischen Ursprungs sind (drei weisen einen B-Haplotyp auf), auch nicht in 13 normalen Chromosomen.

Bei der Mutation S549R wird der stark konservierte Serinrest der Nucleotidbindungsdomäne an der Position 549 zu Arginin umgewandelt (S549 → R); die Codon-änderung ist AGT → AGG. Das CF-Chromosom mit dieser Mutation wird von einem nicht-ashkenasisch/jüdischen Patienten aus Marokko getragen (Familie B). Das Chromosom weist ebenfalls den Haplotyp B auf. Der Nachweis dieser Mutation kann durch ASO-Hybridisation oder Allel-spezifische PCR erbracht werden. Beim ASO-Hybridisationsverfahren wird die genomische DNA zuerst unter Bedingungen von Tabelle 6 durch PCR mit den 11i-5- und 11i-3-Oligonucleotiden amplifiziert; das ASO für die normale Sequenz ist 5'-ACACTGAGTGGAGGTC-3' und das für den Mutanten ist 5'-ACACTGAGGGGAGGTC. Die Oligonucleotidhybridisations-Bedingung ist wie von Kerem et al. (1989), s.o., beschrieben und die Waschungen werden zweimal in 5 x SSC jeweils 10 min lang bei Raumtemperatur durchgeführt, gefolgt von zweimal in 2 x SSC für 30 min, jeweils bei 56°C. Bei der Allel-spezifischen PCR-Amplifikation ist der Oligonucleotidprimer für die normale Sequenz 5'-TGCTCGTTGACCTCCA-3', der



für den Mutanten ist 5'-TGCTCGTTGACCTCCC-3', und der für die gemeinsame, äußere Sequenz ist 11i-5. Die Reaktion wird mit 500 ng genomischer DNA, 100 ng von jeder der Oligonucleotide und 0,5 Einheiten Taq-Polymerase durchgeführt. Das DNA-Templat wird zuerst denaturiert, indem man es 6 min lang bei 94°C erwärmt, gefolgt von 30 Durchgängen bei 94° für 30 s, 55° für 30 s und 72° für 60 s. Die Reaktion wird mit einem 6 min langen Erwärmen bei 72° für 7 min abgeschlossen. Diese S549 → R-Mutation liegt weder in 52 anderen nicht-ΔF508 CF-Chromosomen vor, von denen 11 jüdischen Ursprungs sind (drei weisen einen B-Haplotyp auf), noch in 13 normalen Chromosomen.

10

Bei der S549I-Mutation liegt ein AGT → ATT-Wechsel (Nucleotidpositon 1778 G → T) vor, der die dritte Mutation darstellt, an der dieser Aminosäurecodon beteiligt ist, was zum Verlust der DdeI-Stelle führt. Uns liegt nur ein Beispiel vor, das arabischen Ursprungs ist und sequenziert wurde; kein anderes DdeI-resistentes Chromosom findet sich in 5 anderen arabischen CF, 21 jüdischen CF, 41 kanadischen CF und 13 normalen kanadischen Chromosomen.

15

Bei der Mutation R560T wird das Arginin (AGA) an der Aminosäureposition 560 zu Threonin (ACA) verändert. Das Individuum, das diese Mutation trägt (R560 → T) stammt aus einer Familie, die wir in unserer Arbeit als Familien Nr. 32 identifiziert haben, und das Chromosom wird vom Haplotyp IIIb markiert. Die Mutation erzeugt eine MaeII-stelle, die das PCR-Produkt von Exon 11 (erzeugt mit den Primern 11i-5 und 11i-3 unter Bedingungen von Tabelle 6) in zwei Fragmente mit einer Größe von 214 und 204 bp spaltet. Keines der 36 nicht-ΔF508 CF-Chromosomen (sieben davon weisen den Haplotyp IIIb auf) oder der 23 normalen Chromosomen (16 weisen den Haplotyp IIIb auf) trug diese Sequenzänderung. Die R560 → T-Mutation liegt ebenfalls nicht in acht CF-Chromosomen mit der ΔF508-Mutation vor.

20

25

Bei der Mutation G551D wird Glycin (G) an der Aminosäureposition 551 zu Aspartamsäure (D) umgewandelt. G551 ist ein stark konservierter Rest innerhalb der ATP-Bindungsfaltung. Die entsprechende Codonänderung erfolgt von GGT zu GAT.

30



Die G551 → D-Änderung findet sich in 2 unserer Familien (Nr. 1, Nr. 38) mit Bauchspeicheldrüsen-Unterfunktions- (PI) Patienten und 1 Familie (Nr. 45) mit einem Patienten mit normaler Bauchspeicheldrüsenfunktion (PS). Die anderen CF-Chromosomen in der Familie Nr. 1 und Nr. 38 tragen die ΔF508-Mutation und die der Familie Nr. 54 ist unbekannt. Auf der Grundlage unserer Hypothese der „schweren und milden Mutation“ (Kerem et al., 1989) ist zu erwarten, daß diese Mutation eine „schwere“ ist. Alle 3 Chromosomen, die diese Mutation tragen, gehören zur Gruppe IIIb. Diese G551 → D-Substitution stellt keinen Sequenzpolymorphismus dar, da der Wechsel nicht in 35 anderen CF-Chromosomen ohne die ΔF508-Deletion (5 davon aus Gruppe IIIb) und 19 normalen Chromosomen (einschließlich 5 aus Gruppe IIIb) nachgewiesen wird. Um diese Mutation nachzuweisen, kann die genomische DNA-Region unter Bedingungen von Tabelle 6 durch PCR mit den Primern 11i-5 (5'-CAACTGTGGTTAAAGCAATAGTGT-3') und 11i-3 (5'-GCACAGATTCTGAGTAACCATAAT-3') amplifiziert und auf die Anwesenheit einer MboI (Sau3A) -Stelle untersucht werden, die durch eine Nucleotidwechsel erzeugt wird; die ungeschnittene (normale) Form weist eine Länge von 419 bp auf und die Digestionsprodukte (aus der Mutantenform) sind 241 und 178 bp.

3.8.7 MUTATIONEN IN EXON 12

Bei der Y563N-Mutation wird ein T zu A-Wechsel an der Nucleotidposition 1820 in Exon 12 nachgewiesen. Dieser Wechsel würde eine Änderung von Tyr zu Asn an der Aminosäureposition 563 zur Folge haben. Er findet sich in einer einzigen Familie mit 2 PS-Patienten, aber die Mutation in dem anderen Chromosom ist unbekannt. Wir denken, daß Y563N wahrscheinlich eine Fehlsinnmutation ist, da 81) der Wechsel von T zu A sich nicht in 59 anderen CF-Chromosomen findet, wovon 8 den gleichen Haplotyp (IIa) und 30 ΔF508 aufweisen; und (2) diese Änderung sich nicht in 54 normalen Chromosomen findet, wovon 39 den 11a-Haplotyp aufweisen. Lieder ist die Änderung der Aminsäure nicht drastisch genug, um ein starkes Argument zu liefern. Diese vermutete Mutation kann durch ASO-Hybridisation mit einer normalen (5'-AGCAGTATACAAAGATGC-3') und



einer Mutanten- (5'-AGCAGTAAACAAAGATGC-3') Oligonucleotidsode nachgewiesen werden. Die Waschbedingungen sind 54°C mit 2 x SSC.

Bei der P574H-Mutation wird das C an der Nucleotidposition 1853 zu A umgewandelt. Obwohl die Aminosäure Pro an dieser Position unter verschiedenen ATP-Bindungsfaltungen nicht stark konserviert ist, könnte die Veränderung von C zu His eine drastische Substitution darstellen. Dieser Wechsel wird nicht in 52 anderen CF-Chromosomen und auch nicht in 15 normalen Chromosomen nachgewiesen, von denen 4 den gleichen Haplotyp der Gruppe IV aufweisen. Aufgrund dieser Argumente nehmen wir an, daß es sich bei P574H um eine Mutation handelt. Um diese vermutete Mutation nachzuweisen, muß man die folgenden ASOs verwenden: 5'-GACTCTCATTTTGGGA-3' für den normalen und 5'-GACTCTCATTTTGGGA-3' für den Mutanten. Das Waschen sollte bei 47° in 2 x SSC durchgeführt werden.

3.8.7 (a) MUTATIONEN IN EXON 17 6

Bei der L1077P-Mutation wird das T an der Nucleotidposition 3362 zu C verändert. Dies führt zu einer Änderung der Aminosäure Leu zu Pro an der Amino-position 1077 in Fig. 1. Wie bei den anderen Mutationen in diesem Exon wird die genomische DNA durch Verwendung der Primer von Tabelle 5, nämlich 17bi-5 und 17bi-3, amplifiziert. Die Reaktionsbedingungen bei der Amplifizierung der genomischen DNA sind in Tabelle 6 angegeben.

Die Y1092X-Mutation beinhaltet eine Änderung von C an der Nucleotidposition 3408 zu A. Dies würde zu einem Proteinsynthese-Abbruch an der Amino-position 1092 führen. Daher kommt die Aminosäure Tyr in dem abgeschnittenen Polypeptid nicht vor. Wie bei den vorstehend genannten Verfahren handelt es sich bei den Primern, die bei der Amplifikation dieser Mutation verwendet werden, um 17bi-5 und 17bi-3.



3.8.8 MUTATIONEN IM EXON 19

3659 del C ist eine Frameshift-Mutation im Exon 19 in einem einzigen Chromo-
som (Familie Nr. 2 aus Toronto); Deletion von C an der Nucleotidposition 3659 oder
5 3960; Haplotyp Iia; nicht in 57 nicht- $\Delta R508$ CF-Chromosomen (7 aus Iia) und 50
N-Chromosomen (43 aus Iia) vorhanden; die Deletion kann mit PCR mit einem
gebräuchlichen Oligonucleotidprimer 19i-5 (siehe Tabelle 5) und 2 ASO-Primern,
HSc8 (5'-GTATGGTTTGGTTGACTT GG3') für das normale und
HSC9 (5'-GTATGGTTTGGTTGACTTGT-3') für das Mutantenallel nachgewiesen
10 werden; die PCR-Bedingung ist wie gewöhnlich, abgesehen davon, daß die Erhitzungs-
temperatur bei 60°C liegt, um die Spezifität zu verbessern.

3.8.9 MUTATIONEN IM INTRON 4

15 Bei der 621 + 1G \rightarrow T-Mutation kommt eine einzige bp-Änderung vor, die die
Spleißstelle (GT \rightarrow TT) am 3'-Ende von Exon 4 beeinflusst; diese Mutation wird in
5 frankokanadischen CF-Chromosomen (eine in jeder der Familien Nr. 22, 23, 26, 36
und 53 aus Toronto), aber nicht in 33 anderen CF-Chromosomen (18 aus der gleichen
Gruppe, Gruppe I) und 29 N-Chromosomen (13 aus Gruppe I) nachgewiesen; die
20 Mutation erzeugt eine MseI-Stelle; die genomische DNA kann durch 2 Intronprimer,
4i-5 und 4i-3, amplifiziert und mit MseI zerschnitten werden, um die normalen von den
Mutantenallelen unterscheiden zu können; die normalen ergeben 4 Fragmente mit einer
Größe von 33, 35, 71 und 298 bp; das 298 bp-Fragment im Mutanten wird von dem
Enzym gespalten und ergibt 54 und 244 bp-Fragmente.

25

3.8.10 MUTATIONEN IM INTRON 5

Bei der 711 + 1G \rightarrow T-Mutation tritt dieser Wechsel von G zu T an der Spleiß-
verbindung hinter Exon 5 auf. Diese Mutation findet sich auf dem CF-Chromosom der
30 Mutter der Familie Nr. 22, einer frankokanadischen Familie aus Chicoutimi. Bei der
anderen Mutation in dieser Familie handelt es sich um 621 + 1G \rightarrow T.



3.8.11 MUTATIONEN IM INTRON 10

Bei der 1717-1G → A-Mutation findet sich eine vermutete Spleißmutation vor
5 dem Exon 11. Diese Mutation befindet sich auf dem letzten Nucleotid des Introns vor
dem Exon 11. Die Mutation kann mit den folgenden ASOs nachgewiesen werden:
normal = 5'-TTTGGTAATAGGACATCTCC-3';
Mutant-ASO = 5'-TTTGGTAATAAGACATCTCC-3'. Die Waschbedingungen nach
der Hybridisation sind 5 x SSC zweimal 10 min lang bei Raumtemperatur, 2xSSC
10 zweimal 30 min lang bei 47° für das Mutanten- und 2 x SSC zweimal 30 min lang bei
48° für das normale ASO. Uns liegt nur 1 Beispiel von einem arabischen Patienten vor,
und es gibt keine Haplotypdaten. Die Mutation findet sich nicht in 5 anderen arabischen,
21 jüdischen und 41 kanadischen CF-Chromosomen, auch nicht in 13 normalen
Chromosomen.

15

3.9 DNA-SEQUENZPOLYMORPHISMEN

Nucleotidposition	Aminosäure-Änderung
1540 (A oder G)	Met oder Val
1716 (G oder A)f	keine Änderung (Glu)
2694 (T oder G)	keine Änderung (Thr)
356 (G oder A)	Arg oder Gln

Ein Polymorphismus wird an der Nucleotidposition 1540 nachgewiesen. Der
20 A-Rest kann mit G substituiert werden, was die entsprechende Aminosäure von Met zu
Val ändert. An der Position 2694 kann der T-Rest ein G sein; obwohl dies die kodierte
Aminosäure nicht ändert. Der Polymorphismus kann durch die Restriktionsenzyme
AvaII oder Sau9GI nachgewiesen werden. Diese Änderungen kommen in der normalen
Population vor und zeigen eine gute Korrelation mit den Haplotypen, aber nicht bei CF-
25 Krankheit.



Es kann eine Änderung von G nach A im letzten Nucleotid des Exons 10
(Nucleotidposition 1716) vorhanden sein. Wir denken, daß diese Nucleotidsubstitution
einen Sequenzpolymorphismus darstellt, da (a) sie nicht die Aminosäure verändert,
(b) sie wahrscheinlich keinen Spleißdefekt verursacht und (c) sie auf einigen normalen
5 Chromosomen vorkommt. In zwei kanadischen Familien findet sich dieses seltene Allel
in Assoziation mit dem Haplotyp IIIb.

Man fand, daß das häufiger vorkommende Nucleotid an 356 (G) in dem normalen
Chromosom des Vaters der Familie Nr. 54 zu A verändert wurde. Die Aminosäure ver-
ändert sich von Arg zu Gln.
10

4.0 CFTR-PROTEIN

Wie mit Bezug auf die DNA-Sequenz der Fig. 1 erörtert, sagte die Analyse der
15 Sequenz der sich überlagernden cDNA-Klone ein unreifes Polypeptid aus 1480
Aminosäuren mit einer Molekülmasse von 168138 Dalton voraus. Wie später be-
schrieben, kann aufgrund von Polymorphismen im Protein das Molekulargewicht des
Proteins aufgrund von möglichen Substitutionen oder der Deletion von bestimmten
Aminosäuren variieren. Das Molekulargewicht ändert sich auch aufgrund der Addition
20 von Kohlenhydrateinheiten, wodurch ein Glycoprotein gebildet wird. Selbstverständlich
ähnelt das funktionelle Protein in der Zelle dem unreifen Polypeptid, kann aber auf-
grund von Zellmetabolismen modifiziert werden.

Demgemäß ist ein gereinigtes normales CFTR-Polypeptid durch ein Molekular-
gewicht von etwa 170000 Dalton gekennzeichnet und dadurch, daß es eine trans-
25 membranische ionenleitende Aktivität durch die Epithelzellen aufweist. Das normale
CFTR-Polypeptid, das praktisch frei von anderen menschlichen Proteinen ist, wird von
den vorstehend genannten DNA-Sequenzen und gemäß einer Ausführungsform, der von
Fig. 1, kodiert. Solche in Polypeptid zeigt die immunologische oder biologische
30 Aktivität von normalem CFTR-Polypeptid. Wie später erörtert werden wird, können das
CFTR-Polypeptid und dessen Fragmente durch chemische oder enzymatische Peptid-



synthese hergestellt werden oder in einem geeigneten kultivierten Zellsystem exprimiert werden. Die Erfindung stellt gereinigtes 507-Mutanten-CFTR-Polypeptid bereit, das durch eine mit cystischer Fibrose assoziierten Aktivität in menschlichen Epithelzellen gekennzeichnet ist. Solch ein 507-Mutanten-CFTR-Polypeptid kann, da es praktisch frei von anderen menschlichen Proteinen ist, durch die 507-Mutanten-DNA-Sequenz kodiert werden.

4.1 CFTR-STRUKTUR

Das charakteristischste Merkmal des vorhergesagten Proteins ist die Anwesenheit von zwei sich wiederholenden Motiven, von denen jedes aus einer Gruppe von Aminosäureresten besteht, die in der Lage sind, die Membran mehrmals zu umfassen, gefolgt von sequenzähnlichen Consensusnucleotid (ATP) -Bindungsfalten (NBFs) (Fig. 11, 12 und 15). Diese Eigenschaften sind denen des gegen Kombinationsheilmittel resistenten Säuger-P-Glycoproteins bemerkenswert ähnlich, wodurch impliziert wird, daß das vorausgesagte CF-Genprodukt in den Transportsubstanzen (Ionen) durch die Membran enthalten und wahrscheinlich ein Mitglied der Membranprotein-Superfamilie ist.

Fig. 13 ist ein schematisches Modell des vorausgesagten CFTR-Proteins. In Fig. 13 zeigen Zylinder Membran-umspannende Helices an, die beiden offenen Kreise zeigen NBFs an. Der gestreifte Bereich ist die polare R-Domäne. Die 6 Membran-umspannenden Helices in jeder Hälfte des Moleküls sind als Zylinder dargestellt. Die inneren cytoplasmisch ausgerichteten NBFs als Kreise sind mit Schlitzen dargestellt, um die Eintrittsmöglichkeit des Nucleotids anzuzeigen. Die große polare R-Domäne, die die beiden Hälften verbindet, ist durch den gestreiften Bereich dargestellt. Einzelne geladene Aminosäuren innerhalb der Transmembransegmente und auf der R-Domänenoberfläche sind als kleine Kreise dargestellt, die das Ladungszeichen aufweisen. Die Nettoladungen auf den inneren und äußeren Schleifen, die die Membranzylinder verbinden, und in Regionen der NBFs sind in offenen Vierecken enthalten. Stellen für die Phosphorylation durch die Proteinkinasen A oder C sind durch geschlossene bzw. ge-



geschlossene Dreiecke dargestellt. K, R, H, D und E stellen die Standardnomenklatur für die Aminosäuren, Lysin, Arginin, Histidin, Asparaginsäure bzw. Glutaminsäure dar.

Jede der vorausgesagten Membran-assoziierten Regionen des CFTR-Proteins besteht aus 6 stark hydrophoben Segmenten, die in der Lage sind, eine Lipid-Doppelschicht gemäß den Algorithmen von Kyte und Doolittle und von Garnier et al. (*J. Mol. Biol.* 120, 97 (1978) (Fig. 13) zu umspannen. Auf die Membran-assoziierten Regionen folgt jeweils eine große hydrophile Region, die die NBFs enthält. Aufgrund von vergleichender Sequenzanordnung mit anderen bekannten Nucleotid-Bindungsproteinen, umfaßt jede der vermuteten NBFs in CFTR mindestens 150 Reste (Fig. 13). Die 3 bp Deletion an der Position 507, wie in CF-Patienten nachgewiesen, liegt zwischen den 2 am stärksten konservierten Segmenten des ersten NBF im CFTR. Die Identität der Aminosäuresequenz zwischen der Region, die die Isoleucin-Deletion umgibt und den entsprechenden Regionen einer Reihe anderer Proteine deutet darauf hin, daß diese Region von funktioneller Wichtigkeit ist (Fig. 15). Eine hydrophobe Aminosäure, üblicherweise eine mit einer aromatischen Seitenkette, ist in den meisten dieser Protein an der Position, die I507 auf dem CFTR-Protein entspricht, vorhanden. Selbstverständlich können Aminosäure-Polymorphismen als Ergebnis von DNA-Polymorphismen existieren. Ähnlich dazu deuteten Mutationen an den anderen Positionen im Protein darauf hin, daß die entsprechenden Regionen des Proteins ebenfalls von funktioneller Wichtigkeit sind. Solche zusätzlichen Mutationen schließen die folgenden Substitutionen ein:

- i) Glu für Gly an der Aminosäureposition 85;
- ii) Thr für Ile an der Aminosäureposition 148;
- 25 iii) Arg für Gly an der Aminosäureposition 178;
- iv) Glu für ALA an der Aminosäureposition 455;
- v) Stopcodon für Gln an der Aminosäureposition 493;
- vi) Stopcodon für Gly an der Aminosäureposition 542;
- vii) Arg für Ser oder Ile für Ser an der Aminosäureposition 549;
- 30 viii) Asp für Gly an der Aminosäureposition 551;
- ix) Thr für Arg an der Aminosäureposition 560;



- x) Asn für Tyr an der Aminosäureposition 563;
- xi) His für Pro an der Aminosäureposition 574;
- xii) Pro für Leu an der Aminosäureposition 1077;
- xiii) Stopcodon für Tyr an der Aminosäureposition 1092.

5

Fig. 15 zeigt eine vergleichende Anordnung der 3 am stärksten konservierten Segmente der extendierten NBFs von CFTR mit vergleichbaren Regionen anderer Proteine. Diese 3 Segmente bestehen aus den Resten 433 – 473, 488 – 513 und 542 bis 584 der N-terminalen Hälfte und 1219 – 1259, 1277 – 1302 und 1340 – 1382 der C-terminalen Hälfte von CFTR. Die dicken Striche darüber betonen die Regionen mit der stärksten Ähnlichkeit. Eine zusätzliche allgemeine Homologie ist selbst ohne Einführung von Lücken zu sehen.

Trotz der allgemeinen Symmetrie der Struktur des Proteins und der Sequenzkonservierung der NBFs, ist die Sequenzhomologie zwischen den beiden Hälften des vorhergesagten CFTR-Proteins mäßig. Dies zeigt sich an Fig. 12, wo die Aminosäuren 1-1480 auf jeder Achse dargestellt sind. Linien auf jeder Seite der Identitätsdiagonalen zeigen die Positionen mit inneren Ähnlichkeiten an. Darum sind, obwohl vier Gruppen mit innerer Sequenzidentität unter Verwendung der Dayhoff-Punktmatrix, wie von Lawrence et al. angewendet [C.B. Lawrence, D.A. Goldman und R.T. Hood, *Bull. Math. Biol.* 48, 569 (1986)], nachgewiesen werden können, wie in Fig. 12 gezeigt, drei davon nur als niedrige Schwellenwerte für die Standardabweichung erkennbar. Die stärkste Identität befindet sich zwischen Sequenzen an den Carboxylenden der NBFs. Von den 66 vergleichend angeordneten Resten sind 27 % identisch und weitere 11 % sind in ihrer Funktion ähnlich. Die insgesamt schwache innere Homologie steht im Gegensatz zu dem viel höheren Grad (> 70 %) im P-Glycoprotein, für das eine Genduplikationshypothese aufgestellt wurde (Gros et al., *Cell* 47, 381 (1986), Gerlach et al., *Nature* 324, 485, (1986), Gros et al., *Mol. Cell. Biol.* 8, 2770 (1988)). Die fehlende Konservierung in den relativen Positionen der Exon/Introngrenzen kann gegen ein solches Modell für CFTR sprechen (Fig. 2).



Da sich offensichtlich keine Signalpeptidsequenz an dem Aminoterminus von CFTR befindet, ist das stark geladene hydrophile Segment, das der ersten Transmembransequenz vorausgeht, wahrscheinlich im Cytoplasma ausgerichtet. Jede der zwei Gruppen hydrophober Helices formt wahrscheinlich 3 transverse Schleifen über die Membran und wahrscheinlich ist nur eine geringe Sequenz des gesamten Proteins der äußeren Oberfläche ausgesetzt, außer der Region zwischen dem Transmembransegment 7 und 8. Interessanterweise enthält die letztgenannte Region zwei potentielle Stellen für die N-verknüpfte Glycosylierung.

Jeder der Membran-assoziierten Regionen folgt ein NBF, wie vorstehend angegeben. Darüber hinaus kann eine stark geladene Cytoplasmadomäne in der Mitte des vorausgesagten CFTR-Polypeptids identifiziert werden, die die 2 Hälften des Proteins verbindet. Diese Domäne, als R-Domäne bezeichnet, ist nach seiner Wirkung als ein einziges großes Exon definiert, in dem 69 der 241 Aminosäuren polare Reste sind, die in alternierenden Clustern mit positiver und negativer Ladung angeordnet sind. Darüber hinaus liegen 9 der 10 Consensussequenzen, die für die Phosphorylierung (s.o.) durch die Proteinkinase A (PKA) erforderlich sind, und 7 der potentiellen Substratstellen für die Proteinkinase C (PKC), die sich in CFTR finden, auf diesem Exon.

4.2 FUNKTION DES CFTR

Die Eigenschaften des CFTR können aus dem Vergleich mit anderen Membran-assoziierten Proteinen (Fig. 15) abgeleitet werden. Zusätzlich zur allgemeinen Ähnlichkeit der Struktur mit dem Säuger-P-Glycoprotein, zeigt jede der beiden vorausgesagten Domänen im CFTR auch eine bemerkenswerte Ähnlichkeit mit der Einzeldomänenstruktur des Hemolysins B von *E. coli* und dem Produkt des White-Gens von *Drosophila*. Diese letztgenannten Proteine sind am Transport des lytischen Peptids des Hämolytinsystems bzw. von Augenpigmentmolekülen beteiligt. Das Vitamin B12-Transportsystem von *E. coli*, BtuD und MbpX, bei dem es sich um ein Leberblümchen-Chloroplastgen handelt, dessen Funktion unbekannt ist, weisen ebenfalls ein ähnliches Strukturmotiv auf. Außerdem hat das CFTR-Protein eine strukturelle Ähnlichkeit mit



mehreren der periplasmatischen Lösungs-Transportsysteme von gramnegativen Bakterien, wo die Transmembranregion und die ATP-Bindungsfalten in separaten Proteinen enthalten sind, die in Übereinstimmung mit einem dritten Substrat-bindenden Polypeptid wirken.

5

Die allgemeine strukturelle Anordnung der Transmembrandomänen in CFTR ist der mehrerer Kationenkanalproteine und einiger Kationen-translozierender ATPasen, ebenso wie der kürzlich beschriebenen Adenylatcyclase von Rinderhirn ähnlich. Der funktionelle Stellenwert dieser topologischen Klassifizierung, die aus 6 Transmembran-

10 domänen besteht, bleibt spekulativ.

Kurze Regionen mit Sequenzidentität wurden ebenfalls zwischen den vermuteten Transmembranregionen von CFTR und anderen Membran-umspannenden Proteinen nachgewiesen. Interessanterweise liegen auch Sequenzen vor, 18 Aminosäuren lang, die

15 ungefähr 50 Reste vom Carboxylterminus von CFTR und dem raf Serin/Threonin-Kinase-Protooncogen von *Xenopus laevis* entfernt sind, die in 12 dieser Positionen identisch sind.

Schließlich wurde eine Aminosäuresequenz-Identität (10/13 konservierte Reste)

20 zwischen einem hydrophilen Segment (Position 701-713) innerhalb der hoch aufgeladenen R-Domäne von CFTR und einer Region, die der ersten Transmembranschleife der Natriumkanäle unmittelbar vorangeht, in sowohl Rattenhirn als auch Aal entdeckt. Die geladene R-Domäne von CFTR wird nicht mit dem topologisch nahen P-Glycoprotein geteilt; das 241 Aminosäure-Verbindungspeptid stellt offensichtlich den größten

25 Unterschied zwischen den beiden Proteinen dar.

Zusammenfassen ist zu sagen, daß die Merkmale der primären Struktur des CFTR-Proteins anzeigen, daß es Eigenschaften besitzt, die geeignet sind, an der Regulation und Steuerung des Ionentransports in den Epithelzellen von Geweben, die

30 mit CF befallen sind, teilzunehmen. Eine feste Anheftung an der Membran in zwei Regionen dient dazu, seine drei wichtigsten intracellulären Domänen (die Nucleotid-



Bindungsfalten 1 und 2 und die R-Domäne) in der Nähe der cytoplasmischen Oberfläche der Zellmembran zu positionieren, wo sie die Ionenbewegung durch Kanäle modulieren kann, die entweder durch CFTR-Transmembransegmente an sich oder durch andere Membranproteine gebildet werden.

5

Angesichts der genetischen Daten, der Gewebespezifität und der vorausgesagten Eigenschaften des CFTR-Proteins kann vernünftigerweise geschlossen werden, daß CFTR direkt für CF verantwortlich ist. Es bleibt jedoch unklar, wie CFTR an der Regulation der Ionenleitung durch die apikale Membran der Epithelzellen beteiligt ist.

10

Möglicherweise dient CFTR selbst als Ionenkanal. Wie in Fig. 13 dargestellt, enthalten 10 der 12 Transmembranregionen eine oder mehrere Aminosäuren mit geladenen Seitenketten, eine Eigenschaft, die der der Untereinheiten des Gehir-Natriumkanals und GABA-Rezeptor-Chloridkanals ähnelt, wo geladene Reste in 4 der 6 bzw. 3 der 4 Membran-assoziierten Domänen pro Untereinheit oder Wiederholungseinheit vorliegen. Man nimmt an, daß die amphipathische Natur dieser Transmembransegmente zur Fähigkeit dieser Moleküle, Kanäle zu bilden, beiträgt. Andererseits ist es möglich, daß CFTR kein Ionenkanal ist, sondern dazu dient, die Aktivität der Ionenkanäle zu regulieren. Für die letztgenannte Annahme spricht, daß es sich bei keinem der gereinigten Polypeptide aus Trachea oder Niere, die in der Lage sind, Chloridkanäle in Lipidmembranen wiederherzustellen [Landry et al., *Science* 224:1469 (1989)], um CFTR zu handeln scheint, wenn man von der Molekülmasse ausgeht.

20

In jedem Fall legt die Anwesenheit von ATP-Bindungsdomänen in CFTR nahe, daß die ATP-Hydrolyse direkt an der Transportfunktion beteiligt und dafür notwendig ist. Die hohe Dichte der Phosphorylierungsstellen für PKA und PKC und die Cluster von geladenen Stellen in der R-Domäne können beide dazu dienen, diese Aktivität zu regulieren. Die Deletion eines Phenylalaninrests im NBF kann das korrekte Binden von ATP oder die Konformationsänderung, die dieses normalerweise bewirkt, verhindern, und infolgedessen die beobachtete Unempfindlichkeit gegenüber der Aktivierung durch PKA- oder PKC-vermittelte Phosphorylierung des apikalen Chlorid-Leitungsweges bei

30



CF zur Folge haben. Da das vorausgesagte Protein mehrere Domänen besitzt und zu einer Proteinfamilie gehört, die häufig die Funktion von Teilen von molekularen Multi-komponentensystemen hat, kann das CFTR auch an Epithelgewebefunktionen mit Aktivitäten oder Regulationen, die nicht mit dem Ionentransport in Zusammenhang stehen, beteiligt sein.

Da nun das isolierte CF-Gen c(DNA) zur Verfügung steht, ist es möglich, den biochemischen Grunddefekt bei CF zu definieren, und die Steuerung der Ionentransportwege in Epithelzellen im allgemeinen weiter zu erklären. Am wichtigsten ist, daß das Wissen, das bisher über die vorausgesagte Struktur des CFTR erlangt wurde, zusammen mit den zusätzlichen Informationen aus Studien des Proteins selbst die Grundlage für die Entwicklung verbesserter Behandlungsmethoden der Krankheit liefert. In solchen Studien wurden Antikörper zum CFTR-Protein entwickelt, wie später beschrieben wird.

5.0 CF-SCREENING

5.1 DIAGNOSE AUF DNA-GRUNDLAGE

Mit dem Wissen um die Mutationen an der 85-, 148-, 178-, 455-, 493-, 507-, 542-, 549-, 551-, 560-, 563-, 574-, 1077- und 1092-Aminosäureposition und den Nucleotidsäuresequenz-Varianten an den DNA-Sequenzpositionen 129, 556, 621+1, 711+1, 1717-1 und 3659 wie hierin offenbart, kann ein Träger-Screening und eine pränatale Diagnose wie folgt durchgeführt werden können.

Die Population mit einem hohen Risiko für cystische Fibrose ist die kaukasische. Beispielsweise würde jede kaukasische Frau und/oder jeder kaukasische Mann im fortpflanzungsfähigen Alter gescreent werden, um zu bestimmen, ob sie oder er ein Träger ist (eine ungefähr 5%-ige Wahrscheinlichkeit in jedem Individuum). Wenn beide Träger sind, sind sie ein Paar, bei dem das Risiko eines an cystischer Fibrose erkrankten Kindes besteht. Jedes Kind des risikobehafteten Paares weist eine 25%-ige Wahrscheinlichkeit auf, an cystischer Fibrose zu erkranken. Das Verfahren zur Bestimmung des Trägerstatus unter Verwendung der hierin offenbarten Sonden ist wie folgt.



Um der Kürze willen wird die Erörterung des Screenings mittels einer der selektierten Mutationen auf die I507-Mutation konzentriert. Selbstverständlich kann das Screening auch mittels einer der anderen Mutationen oder mittels mehrerer Mutationen
5 in einem Screening-Verfahren oder einem Mutations-Nachweisverfahren dieses Abschnitts über das CF-Screenen durchgeführt werden, die die DNA-Diagnose und den Mutationsnachweis beinhalten.

Ein Hauptanwendungsgebiet der DNA-Sequenzinformation über das normale und
10 das 507-Mutanten-CF-Gen liegt auf dem Gebiet des genetischen Testens, des Trägernachweises und der pränatalen Diagnose. Individuen, die Mutationen im CF-Gen tragen (Krankheitsträger oder Patienten), können auf DNA-Level unter Verwendung einer Reihe von Techniken nachgewiesen werden. Die genomische DNA, die für die Diagnose verwendet wird, kann aus Körperzellen, wie denen, die im peripheren Blut, im
15 Urin und im Speichel, aus der Gewebebiopsie, in chirurgischen Gewebeproben und aus Autopsiematerial vorkommen, erhalten werden. Die DNA kann direkt für den Nachweis einer spezifischen Sequenz verwendet werden oder kann enzymatisch *in vitro* mittels PCR [Saiki et al., *Science* 230: 1350-1353 (1985), Saiki et al., *Nature* 324: 163-166 (1986)] vor der Analyse amplifiziert werden. RNA oder ihre cDNA-Form kann auch für
20 den gleichen Zweck verwendet werden. Kürzliche Besprechungen dieses Themas wurden von Caskey [*Science* 236:1223-8 (1989) und von Landegren et al. (*Science* 242: 229-237 (1989))].

Der Nachweis einer spezifischen DNA-Sequenz kann durch Methoden wie die
25 Hybridisation unter Verwendung spezifischer Oligonucleotide [Wallace et al., *Cold Spring Harbour Symp. Quant. Biol.* 51: 257-261 (1986)], direktes DNA-Sequenzieren [Church und Gilbert, *Proc. Nat. Acad. Sci. U.S.A.* 81: 1991-1995 (1988)], Verwendung von Restriktionsenzymen [Flavell et al., *Cell* 15: 25 (1978), Geever et al., *Proc. Nat. Acad. Sci. U.S.A.* 78: 5081 (1981)], Diskrimination aufgrund von elektrophoretischer
30 Mobilität in Gelen mit denaturierendem Reagens (Myers und Maniatis, *Cold Spring Harbour Symp. Quant. Biol.* 51: 275-284 (1986)), RNase-Schutz (Myers, R.M., Larin, J.,



und T. Maniatis *Science* 230: 1242 (1985)), chemische Abspaltung (Cotton et al., *Proc. Nat. Acad. Sci. U.S.A.* 85: 4397-4401, (1985)) und das Ligase-vermittelte Nachweisverfahren [Landegren et al., *Science* 241: 1077 (1988)] geführt werden.

5 Oligonucleotide, die für normale oder Mutantensequenzen spezifisch sind, werden chemisch mittels im Handel erhältlicher Maschinen synthetisiert, radioaktiv mit Isotopen (wie ^{32}P) oder nicht-radioaktiv (mit Markierungen wie Biotin (Ward und Langer et al., *Proc. Nat. Acad. Sci. U.S.A.* 78: 6633-6657 (1981)) markiert und zu individuellen DNA-Markern, die auf Membranen oder auf anderen festen Unterlagen
10 fixiert wurden, durch Dot-Blot oder Übertragung von Gelen nach der Elektrophorese hybridisiert. Die Anwesenheit oder Abwesenheit dieser spezifischen Sequenzen werden durch Methoden wie die Autoradiographie oder fluorometrische (Landegren et al., 1989, s.o.) oder colorimetrische Reaktionen (Gebeyehu et al., *Nucleic Acids Research* 15: 4513-4534 (1987)) sichtbar gemacht. Eine Ausführungsform dieses Oligonucleotid-
15 Screeningverfahrens wurde bei dem Nachweis der I507-Deletion, wie hierin beschrieben, angewendet.

Sequenzunterschiede zwischen normalen und Mutanten können durch das direkte DNA-Sequenzierungsverfahren von Church und Gilbert (s.o.) gezeigt werden. Klonierte
20 DNA-Segmente können als Sonden verwendet werden, um spezifische DNA-Segmente nachzuweisen. Die Empfindlichkeit dieses Verfahrens wird stark verbessert, wenn es mit 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)]. Im letztgenannten Verfahren wird ein Sequenzierungsprimer, der innerhalb der amplifizierten Sequenz
25 liegt, mit doppelsträngigem PCR-Produkt oder einsträngiger Schablone verwendet, die durch modifizierte PCR erzeugt wurden. Der Sequenznachweis wird durch herkömmliche Verfahren mit radiomarkierten Nucleotiden oder durch automatische Sequenzierverfahren mit Fluoreszenzmarkern durchgeführt.

30 Sequenzänderungen können gelegentlich zufällige Restriktionsenzym-Erkennungsstellen erzeugen, die durch Verwendung der geeigneten Enzymdigestion, gefolgt



von herkömmlicher Gel-Blotthybridisation (Southern, *J. Mol. Biol.* 98: 503 (1975)), sichtbar werden. DNA-Fragmente, die die Stelle tragen (entweder normal oder mutant) werden durch ihre reduzierte Größe oder die Erhöhung ihrer entsprechenden Restriktionsfragmentzahlen nachgewiesen. Genomische DNA-Proben können auch durch PCR
5 vor der Behandlung mit dem geeigneten Restriktionsenzym amplifiziert werden; Fragmente mit unterschiedlichen Größen werden dann unter UV-Licht in Anwesenheit von Ethidiumbromid nach der Gelelektrophorese sichtbar gemacht.

Das genetische Testen aufgrund von DNA-Sequenzunterschieden kann durch
10 Nachweis einer Veränderung der elektrophoretischen Mobilität von DNA-Fragmenten in Gelen mit oder ohne denaturierendes Reagens durchgeführt werden. Kleine Sequenzdeletionen und -insertionen können durch hochauflösende Gelelektrophorese sichtbar gemacht werden. Beispielsweise ist das PCR-Produkt mit der 3 bp Deletion klar von der normalen Sequenz auf einem 8%-igen nicht-denaturierenden Polyacrylamidgel zu unterscheiden. DNA-Fragmente mit unterschiedlichen Sequenzzusammensetzungen können
15 auf denaturierendem Formamidgradientengel, bei dem die Mobilitäten verschiedener DNA-Fragmente im Gel bei verschiedenen Positionen gemäß ihrer spezifischen "Teilschmelz"-Temperaturen (Myers, s.o.) abgebremst werden, unterschieden werden. Darüber hinaus können Sequenzänderungen, insbesondere kleine Deletionen, als Veränderungen im Migrationsmuster von DNA-Heteroduplices bei der nicht-denaturierenden Gelelektrophorese, wie für die 3 bp (I507) -Mutation nachgewiesen wurde, und in
20 anderen experimentellen Systemen [Nagamine et al., *Am. J. Hum. Genet.*, 45:337-339 (1989)] nachgewiesen werden. Andererseits könnte ein Verfahren zum Nachweis einer Mutation, die eine Einbasensubstitution oder andere kleine Veränderung umfaßt, auf der
25 unterschiedlichen Primerlänge in einer PCR basieren. Beispielsweise könnte ein unveränderlicher Primer zusätzlich zu einem Primer, der für eine Mutation spezifisch ist, verwendet werden. Die PCR-Produkte der normalen und Mutantengene können dann differenzierend in Acrylamidgelen nachgewiesen werden.

30 Sequenzänderungen an spezifischen Stellen können auch durch Nucleaseschutz-Assays, wie RNase (Myers, s.o.) und S1-Schutz (Berk, A.J., und P.A. Sahrpe, *Proc.*



Nat. Acad. Sci. U.S.A. 75: 1274 (1978)), das chemische Spaltungsverfahren (Cotton, s.o.) oder das Ligase-vermittelte Nachweisverfahren (Landegren, s.o.) gezeigt werden.

Zusätzlich zu den herkömmlichen Gelelektrophorese- und Blot-Hybridisations-
5 verfahren können DNA-Fragmente auch durch Verfahren sichtbar gemacht werden, bei
denen die einzelnen DNA-Proben nicht auf Membranen fixiert werden. Es können so-
wohl die Sonden- als auch die Zielsequenzen in Lösung vorliegen, oder die Sonden-
sequenz kann fixiert werden [Saiki et al., *Proc. Natl. Acad. Sci. U.S.A.* (86:6230-6234
(1989)]. Eine Reihe von Nachweisverfahren, wie die Autoradiographie unter Ein-
10 beziehung von Radioisotopen, der direkte Nachweis radioaktiven Verfalls (in Anwesen-
heit oder Abwesenheit eines Scintillationsmittels), die Spektrophotometrie einschließ-
lich colorigenischer Reaktionen und die Fluorometrie einschließlich fluorogenischer
Reaktionen können verwendet werden, um spezifische individuelle Genotypen zu
identifizieren.

15

Da mehr als eine Mutation im CF-Gen vermutet wird, wie I507 und F508, stellt
ein multiples System ein ideales Protokoll für das Screening von CF-Trägern und den
Nachweis von spezifischen Mutationen dar. Beispielsweise kann eine PCR mit mehr-
fachen spezifischen Oligonucleotidprimern und Hybridisationssonden verwendet
20 werden, um alle möglichen Mutationen gleichzeitig zu identifizieren (Chamberlain et
al., *Nucleic Acids Research* 16: 1141-1155 (1988)). Das Verfahren kann fixierte
Sequenz-spezifische Oligonucleotid-Sonden beinhalten (Saiki et al., s.o.).

5.2 NACHWEIS DER CF 507-MUTATION

25

Diese Nachweisverfahren können auf die pränatale Diagnose unter Verwendung
von amniotischen Fluidzellen, Chorionzottenbiopsie oder Entmischung fetaler Zellen
vom mütterlichen Kreislauf angewendet werden. Der Test für die CF-Träger in der
Population kann als wesentlicher Bestandteil in ein Gentestprogramm großen Maßstabs
30 für häufige Krankheiten einbezogen werden.



Gemäß einer Ausführungsform der Erfindung kann der Abschnitt des DNA-Segments, das Informationen über eine Mutation liefert, wie die Mutation gemäß dieser Ausführungsform, d.h. der Abschnitt, der die I507-Deletion unmittelbar umgibt, unter Verwendung von Standard-PCR-Verfahren amplifiziert werden [wie von Landegren, Ulf, Robert Kaiser, C. Thomas Caskey und Leroy Hood, DNA Diagnostics – Molecular Techniques and Automation, in *Science* 242: 229-237 (1988)]. Man hält es für möglich, daß es sich bei dem Abschnitt des DNA-Segments, der verwendet wird, um ein einzelnes DNA-Fragment oder eine Mischung verschiedener DNA-Segmente handeln kann. Eine detaillierte Beschreibung dieses Verfahrens folgt.

10

Eine spezifische Region genomischer DNA von der Person oder dem Fetus muß gescreent werden. Solch eine spezifische Region wird durch die Oligonucleotidprimer C16B (5'-GTTTTCTGGATTATGCCTGGCAC-3') und C16D (5'-GTTGGCATGCTTTGATGACGCTTC-3') oder, wie in Fig. 18 gezeigt, von den Primern 10i-5 und 10i-3 definiert. Die spezifischen Regionen unter Verwendung von 10i-5 und 10i-3 wurden durch die Polymerase-Chainreaction (PCR) amplifiziert. 200 bis 400 ng genomischer DNA, entweder von kultivierten Lymphoblasten oder peripheren Blutproben von CF-Individuen und ihren Eltern wurden in jeder PCR mit den vorstehend angegebenen Oligonucleotidprimern verwendet. Die Oligonucleotide wurden mit Oligonucleotid Purification Cartridges™ (Applied Biosystems) oder NENSORB™ PREP Säulen (Dupont) mit Verfahren, die vom Hersteller empfohlen werden, gereinigt. Die Primer wurden für die Extension in einem automatischen Perkin-Elmer/Cetus Thermocycler mit einem Step-Cycle Programm (Transitionseinstellung bei 1.5 min), in 30 Durchgängen mit einem Enddurchgang von 7 min, 30 s lang bei 55°C erhitzt, 60 s lang bei 72°C extendiert (mit 2 Einheiten Taq DNA Polymerase) und 60 s lang bei 94°C denaturiert. Abschnitte des PCR-Produkts wurden gemäß Standardverfahren durch Elektrophorese auf 1,4%-igem Agarosegelen separiert und auf Zeta-bind™-Membran (Biorad) transferiert.

20

25

30

Die normalen und Δ I507 Oligonucleotidsonden von Fig. 19 (jeweils 10 ng) wurden separat mit 10 Einheiten T4 Polynucleotidkinase (Pharmacia) in einer 10 μ l



Reaktion, die 50 mM Tris-HCl (pH 7,6), 10 mM MgCl₂, 0,5 mM Dithiothreitol, 10 mM Spermidin, 1 mM EDTA und 30 – 40 µCi von γ [³²P] – ATP enthielt, 20 – 30 min bei 37°C markiert. Die nicht-eingebauten Radionucleotide wurden mit einer Sphadex G-25-Säule vor der Verwendung entfernt. Die Hybridisationsbedingungen waren wie bereits beschrieben (J.M. Rommens et al., *Am. J. Hum. Genet.* 43, 645 (1988)), abgesehen davon, daß die Temperatur 37°C betragen kann. Die Membrane wurden zweimal bei Raumtemperatur mit 5 x SSC und zweimal bei 39°C mit 2 x SSC gewaschen (1 x SSC = 150 mM NaCl und 15 mM Na-Citrat). Die Autoradiographie wird über Nacht bei Raumtemperatur durchgeführt. Autoradiographen werden entwickelt, um die Hybridisationsergebnisse genomischer DNA mit den 2 spezifischen Oligonucleotidsonden zu zeigen. Sonde C normal weist die normale DNA-Sequenz nach und Sonde C Δ I507 weist die Mutantensequenz nach.

Die genomische DNA-Probe jedes Familienmitglieds kann, wie erklärt, durch die Polymerase-Chainreaction unter Verwendung der Intronsequenzen von Fig. 18 amplifiziert werden, und die Produkte können durch Elektrophorese auf einem 1,4%-igen Agarosegel separiert werden dann auf Zetabind (Biorad) –Membran gemäß Standardverfahren transferiert werden. Die 3 bp Deletion an Δ F507 kann durch ein sehr leicht durchzuführendes Polyacrylamidgel-Elektrophoreseverfahren gezeigt werden. Wenn die PCR-Produkte, die mit den vorstehend genannten 10i-5 und 10i-3 Primern erzeugt wurden, auf ein 5%-iges Polyacrylamidgel aufgetragen werden, 3 h lang bei 20V/cm in einem 90 mM Trisborat-Puffer (pH 8,3) elektrophoretisiert werden, sind DNA-Fragmente mit unterschiedlicher Mobilität für Individuen ohne die 3 bp Deletion, heterozygot oder homozygot für die Deletion klar nachweisbar.

25

Wie bereits mit Bezug auf Fig. 20 erklärt, kann die PCR-amplifizierte genomische DNA einer Gelelektrophorese unterzogen werden, um die 3 bp Deletion zu identifizieren. Wie in Fig. 20 gezeigt, handelt es sich bei den vier Straßen bei Straße 1 um eine Kontrolle mit einer normalen/ Δ I507 Deletion. Die nächste Straße ist der Vater mit einer normalen/ Δ I507 Deletion. Die dritte Straße ist die Mutter mit einer normalen/(Δ F508 Deletion und die vierte Straße ist das Kind mit einer Δ F508/ Δ I507-Deletion. Die Homo-

30



5
10
duplicates erscheinen als feste Banden über die Basis jeder Straße. In den Straßen 1 und 3 zeigen sich die beiden Heteroduplicates sehr klar als zwei voneinander getrennte Banden. In Straße 2 zeigt sich die $\Delta I507$ Mutation des Vaters sehr deutlich, wogegen in der vierten Straße, dem Kind mit den benachbarten 507, 508 Mutationen, keine unterscheidbaren Heteroduplicates vorhanden sind. Somit zeigen sie sich an der Homoduplexlinie. Da der Vater in Straße 2 und die Mutter in Straße 3 Heteroduplexbanden zeigen und das Kind dies nicht tut, zeigt dies, daß das Kind entweder normal oder ein Patient ist. Dies kann weiter untersucht werden, falls nötig, wie durch Embryonenanalyse durch Mischen der 507- und 508-Sonden, um das Vorhandensein von $\Delta I507$ und $\Delta F508$ Mutationen nachzuweisen.

15
Eine ähnliche Veränderung der Gelmobilität für Heteroduplicates, die während PCR geformt wurden, wurde ebenfalls bei experimentellen Systemen berichtet, wo kleine Deletionen beteiligt sind (Nagamine et al., s.o.). Diese Mobilitätsverschiebungen können allgemein als Grundlage für die nicht-radioaktiven genetischen Screeningtests verwendet werden.

5.3 CF-SCREENINGPROGRAMME

20
25
30
Man ist sich darüber im Klaren, daß ungefähr 1 % der Träger mittels der spezifischen $\Delta I507$ Sonden dieser speziellen Ausführungsform der Erfindung nachgewiesen werden können. Somit kann, wenn es sich bei einem getesteten Individuum unter Verwendung der $\Delta I507$ Sonden nicht um einen Träger handelt, dessen Trägerstatus nicht ausgeschlossen werden, es kann irgendeine andere Mutation tragen, wie $\Delta F508$, wie zuvor angegeben. Wenn jedoch sowohl das Individuum als auch der/die Partner(in) des getesteten Individuums Träger der $\Delta I507$ Mutation sind, kann mit Sicherheit ausgesagt werden, daß sie in Risikopaar bilden. Die Sequenz des Gens wie hierin offenbart, ist eine wesentliche Voraussetzung für die Bestimmung der anderen Mutationen.



Die pränatale Diagnose ist eine logische Erweiterung des Träger-Screenings. Es auf zwei verschiedene Arten festgestellt werden, daß ein Paar dem Risiko unterliegt, ein cystische Fibrose-Kind zu bekommen: wenn sie bereits ein cystische Fibrose-Kind haben, sind sie beide, *per definitionem*, obligate Träger des CFTR-Defektgens, und jedes folgende Kind hat eine 25%-ige Wahrscheinlichkeit, an cystischer Fibrose zu erkranken. Ein Hauptvorteil dieser Erfindung eliminiert die Notwendigkeit der Stammbaum-Analyse einer Familie, wogegen gemäß dieser Erfindung, ein Genmutations-Screeningprogramm wie vorstehend geschildert oder ein anderes ähnliches Verfahren angewandt werden kann, um eine genetische Mutation zu identifizieren, die zu einem Protein mit veränderter Funktion führt. Dies hängt nicht von einer früheren Bestätigung der Familie durch ein erkranktes Kind ab. Fetale DNA-Proben können beispielsweise, wie zuvor erwähnt, aus Fruchtwasserzellen und Chorionzotten-Proben erhalten werden. Die Amplifikation mittels Standard-PCR-Verfahren kann dann mit dieser DNA-Schablone durchgeführt werden.

15

Wenn es sich zeigt, daß beide Elternteile Träger mit der $\Delta I507$ -Deletion sind, wäre die Interpretation der Ergebnisse die folgende. Wenn eine Hybridisation der fetalen DNA an die normale Sonde vorliegt, ist der Fetus nicht an cystischer Fibrose erkrankt, obwohl er ein CF-Träger sein kann (50%-ige Wahrscheinlichkeit für jeden Fetus eines Risikopaars). Wenn die fetale DNA nur an die Sonde mit der $\Delta I507$ Deletion hybridisiert und nicht an die normale Sonde, ist der Fetus an cystischer Fibrose erkrankt.

Man ist sich dessen bewußt, daß für diese und andere Mutationen im CF-Gen eine Reihe verschiedener spezifischer Verfahren angewendet werden können, um eine vollständige Diagnose aller potentiellen CF-Träger oder -Patienten bereitzustellen. Eine vollständige Beschreibung dieser Verfahren ist später gegeben.

Die Erfindung stellt somit ein Verfahren und ein Kit zur Bestimmung, ob ein Subjekt ein CF-Träger oder ein CF-Patient ist, bereit. Insgesamt umfaßt das Screening-Verfahren die folgenden Schritte:



Bereitstellen einer biologischen Probe des Subjekts, das gescreent werden soll;
und Bereitstellen eines Assays zum Nachweisen des Vorhandenseins von mindestens
entweder einem 507 CF-Mutantengen, von 507 CF-Mutantengenprodukten oder
5 Mischungen davon in der biologischen Probe.

Das Verfahren kann weiter charakterisiert werden, indem man mindestens eine
weitere Nucleotidprobe einbezieht, bei der es sich um ein anderes DNA-Sequenz-
fragment beispielsweise der DNA der Fig. 1 oder ein anderes DNA-Sequenzfragment
10 des menschlichen Chromosoms 7 handelt, und das an jeder Seite der DNA-Sequenz von
Fig. 1 liegt. In diesem Hinsicht sind die DNA-Fragmente der Intronabschnitte der Fig. 2
für die weitere Bestätigung des Vorhandenseins einer Mutation geeignet. Einzigartige
Aspekte der Introns an den Exongrenzen können bei Screening-Verfahren herangezogen
werden, um das Vorhandensein der Mutation an der I507 Position oder anderen
15 Mutantenpositionen zu bestätigen.

Ein Kit gemäß einer Ausführungsform der Erfindung, der für die Anwendung
beim Screening-Verfahren und für die Durchführung eines Assays auf das Vorhanden-
sein des CF-Mutantengens durch einen Immunoassay geeignet ist, umfaßt:

- 20 (a) einen Antikörper, der spezifisch an ein Genprodukt des mutierten CF-Gens bindet,
das eine Mutation an einer der Positionen 85, 148, 178, 455, 493, 507, 542, 549,
551, 560, 563, 574, 1077 und 1092 aufweist;
- (b) Reagentien zum Nachweisen der Bindung des Antikörpers an das Genprodukt und
- (c) wobei der Antikörper und das Reagens beide in Mengen vorhanden sind, die wirk-
25 sam sind, um den Immunoassay durchzuführen.

Der Kit für die Durchführung eines Assays auf das Vorhandensein des mutierten
CF-Gens kann auch durch Hybridisationsverfahren bereitgestellt werden. Der Kit
umfaßt:

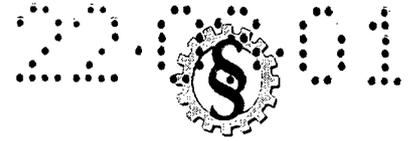


- (a) eine Oligonucleotidsonde, die spezifisch an das mutierte CF-Gen bindet, das eine Mutation an einer der Positionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092 aufweist;
- (b) Reagentien zum Nachweisen der Hybridisation der Oligonucleotidsonde an dem mutierten CF-Gen; und
- (c) wobei die Sonde und die Reagentien beide in Mengen vorhanden sind, die wirksam sind, um den Hybridisationsassay durchzuführen.

5.3 ANTIKÖRPER ZUM NACHWEISEN VON MUTIERTER CFTR

Wie erwähnt werden Antikörper auf Epitope innerhalb des mutierten CFTR-Proteins an den Positionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092 erzeugt, um umfassende Informationen über die Eigenschaften des mutierten Proteins und andere wichtige Informationen zu erhalten, was folgendes einschließt:

1. Die Antikörper können verwendet werden, um weitere Verfahren zum Nachweisen einer der anderen CF-Mutationen bereitzustellen, die in der Synthese eines Proteins mit einer verändertes Stelle resultieren.
2. Antikörper für verschiedene Domänen des mutierten Proteins können verwendet werden, um die topologische Anordnung des Proteins in der Zellmembran zu bestimmen. Dies liefert Informationen über Segmente des Proteins, die für extern zugefügte Modulatoren für die Zwecke der Arzneimitteltherapie zugänglich sind.
3. Die Struktur/Funktions-Beziehungen der Abschnitte des Proteins können mittels spezifischer Antikörper untersucht werden. Beispielsweise ist es möglich, Antikörper in Zellen einzuführen, die jede der geladenen Cytoplasmaschleifen erkennen, die die Transmembransequenzen verbinden, ebenso wie Abschnitte der Nucleotid-Bindungsfalten und der R-Domäne. Der Einfluß dieser Antikörper auf funktionelle Parameter des Proteins liefert Einsichten in den Regulationsmechanismus der Zelle und kann möglicherweise auf Methoden hinweisen, um die Aktivität des defektiven Proteins in einem CF-Patienten zu modulieren.



4. Antikörper mit der geeigneten Avidität ermöglichen auch die Immunpräzipitation und die Immunoaffinitäts-Reinigung des Proteins. Die Immunopräzipitation erleichtert die Charakterisierung der Synthese und der posttranslationalen Modifikation einschließlich der ATP-Bindung und Phosphorylierung. Die Reinigung ist für
5 Untersuchungen der Proteinstruktur und für die Wiederherstellung seiner Funktion notwendig, ebenso wie für die Therapie auf Proteinbasis.

Um die Antikörper herzustellen, können Fusionsproteine, die definierte Abschnitte eines beliebigen der mutierten CFTR-Polypeptide enthalten, in Bakterien durch
10 Expression entsprechender mutierter DNA-Sequenz in einem geeigneten Klonierungsmedium synthetisiert werden. Kleinere Peptide können chemisch synthetisiert werden. Die Fusionsproteine können gereinigt werden, beispielsweise durch Affinitätschromatographie auf Glutathionagarose und die Peptide können an ein Trägerprotein gekoppelt werden (Hämocyanin), mit Freund'schem Adjuvans gemischt und in Kaninchen injiziert
15 werden. Im Anschluß an Auffrischimpfungen in zweiwöchigen Intervallen, werden die Kaninchen ausgeblutet und das Serum wird isoliert. Die entwickelten polyklonalen Antikörper im Serum können dann mit den Fusionsproteinen kombiniert werden. Dann werden Immunoblots durch Anfärben mit beispielsweise Alkalinphosphatase-konjugiertem zweiten Antikörper gemäß dem Verfahren von Blake et al., *Anal. Biochem.*
20 136:175 (1984) gebildet.

Somit ist es möglich, polyklonale Antikörper zu entwickeln, die sowohl für Fusionsprotein, die Abschnitte des mutierten CFTR-Proteins enthalten, als auch für Peptide, die kurzen Segmenten von dessen Sequenz entsprechen, spezifisch sind. Auf
25 ähnliche Weise können Mäuse mit KLH-Konjugaten von Peptiden injiziert werden, um die Produktion von monoklonalen Antikörpern für entsprechende Segmente des mutierten CFTR-Proteins zu initiieren.

Was die Erzeugung monoklonaler Antikörper betrifft, so handelt es sich bei
30 Immunogenen für die Entwicklung monoklonaler Antikörper (mAbs) für das mutierte CFTR-Protein um bakterielle Fusionsproteine [Smith et al., *Gene* 67: 31 (1988)], die



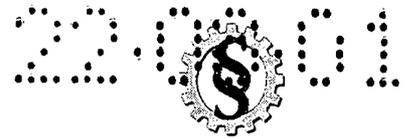
Abschnitte des CFTR-Polypeptids oder von synthetischen Peptiden enthalten, die kurzen (12 bis 25 Aminosäuren langen) Segmenten der mutierten Sequenz entsprechen. Die wesentliche Methodologie ist diejenige von Kohler und Milstein [*Nature* 256: 495 (1975)].

5

Balb/c-Mäuse werden durch intraperitoneale Injektion mit 500 µm reinem Fusionsprotein oder synthetischem Peptid in unvollständigem Freund'schen Adjuvans immunisiert. Eine zweite Injektion wird nach 14 Tagen, eine dritte nach 21 Tagen und eine vierte nach 28 Tagen verabreicht. Die so immunisierten Tiere werden eine, zwei
10 und vier Wochen nach der letzten Injektion geopfert. Die Milz wird entfernt, ihre Zellen werden dissoziiert, gesammelt und mit Sp₂/O-Ag₁₄-Myelomazellen fusioniert gemäß Gefer et al., *Somatic Cell Genetics* 3:231 (1977). Die Fusionsmischung wird in Kulturmedium verteilt, das für die Vermehrung fusionierter Zellen selektiv ist, die gezüchtet werden, bis sie zu etwa 25 % konfluent sind. Zu diesem Zeitpunkt werden Kultur-
15 überstände auf die Anwesenheit von Antikörpern getestet, die mit einem bestimmten CFTR-Antigen reagieren. Ein zweiter mit alkalischer Phosphatase markierte Anti-Maus-Antikörper wird dann für den Nachweis von Positiven verwendet. Zellen aus positiven Kulturschalen werden dann in Kultur vergrößert, ihre Überstände werden für das weitere Testen gesammelt und die Zellen werden tiefgefroren in Cryoschutzmedium
20 gelagert. Um große Mengen an mAB zu erhalten, werden produktive Zellen in das Peritoneum zu 5 x 10⁶ Zellen pro Tier injiziert, und es wird Ascitesflüssigkeit erhalten. Die Reinigung wird durch die Chromatographie auf Protein G- oder ProteinA-Agarose gemäß Ey et al., *Immunochemistry* 15: 429 (1977) durchgeführt.

25 Die Reaktivität dieser mAbs mit dem mutierten CFTR-Protein kann durch Polyacrylamid-Gelelektrophorese von Membranen, die aus Epithelzellen, in denen sie exprimiert und immuno-geblottet wurden, bestätigt werden [Towbin et al., *Proc. Natl. Acad. Sci. U.S.A.* 76: 4350 (1979)].

30 Zusätzlich zur Verwendung von monoklonalen Antikörpern, die für die jeweilige Mutantendomäne des CFTR-Proteins spezifisch sind, um ihre individuellen Funktionen



zu untersuchen, werden andere mABs, die zwischen den normalen und den mutierten Formen des CFTR-Proteins unterscheiden können, verwendet, um das mutierte Protein in Epithelzellproben nachzuweisen, die von Patienten erhalten wurden, beispielsweise von Nasenschleimhaut-Biopsie-“Abstrichen” [R. de-Lough und J. Rutland, *J. Clin. Pathol.* 42, 613 (1989)] oder Hautbiopsieproben, die Schweißdrüsen enthalten.

Antikörper, die zu dieser Unterscheidung fähig sind, werden durch das differenzielle Screening von Hybridomas aus Mäusepaargruppen gewonnen, die mit einem Peptid, das beispielsweise das Isoleucin an der Aminosäureposition 507 enthielt (z.B. GTIKENIIFGVSY), oder einem Peptid, das abgesehen von der Abwesenheit von I507 damit identisch war (GTIKENIFGVSY), immunisiert worden waren. mAbs, die in der Lage sind, die anderen mutierten Formen des CFTR-Proteins in Patienten zusätzlich zu oder anstelle von der I507 Deletion zu erkennen, werden unter Verwendung ähnlicher monoklonaler Antikörper-Produktionsstrategien erhalten.

Antikörper für normale und CF-Versionen des CFTR-Proteins und dessen Segmente werden in der diagnostischen immunocytochemischen und Immunfluoreszenz-Hellfeldmikroskopie verwendet, um die Gewebe-, Zell- und subzelluläre Verteilung von CFTR innerhalb der Organe von CF-Patienten, -Trägern und nicht von CF betroffenen Individuen zu zeigen.

Antikörper werden verwendet, um die Aktivität des CFTR-Proteins in CF-Patienten und in den Zellen von CF-Patienten durch Promoting therapeutisch zu modulieren. Mögliche Methoden für solch eine Modulation können die Stimulation aufgrund der Vernetzung von CFTR-Proteinmolekülen mit mehrwertigen Antikörpern in Analogie zur Stimulation einiger Zelloberflächenmembran-Rezeptoren, wie des Insulinrezeptors [O'Brien et al., *Euro. Mol. Biol. Organ. J.* 6: 4003 (1987)], des epidermalen Wachstumsfaktor-Rezeptors [Schreiber et al., *J. Biol. Chem.* 258:846 (1983)] und von T-Zellenrezeptor-assoziierten Molekülen, wie CD4 [Veillette et al., *Nature* 338: 257 (1989)], beinhalten.



Antikörper werden verwendet, um die Zufuhr von therapeutischen Mitteln auf die Zellen, die defektives CF-Protein in CF exprimieren, zu richten. Zu diesem Zweck werden die Antikörper in einen Träger, beispielsweise Liposom [Mattay et al., *Cancer Res.* 46:4904 81986)] eingebunden, der das therapeutische Agens, beispielsweise ein
5 Arzneimittel oder das normale Gen, trägt.

5.5 RFLP-ANALYSE

Die DNA-Diagnose wird gegenwärtig angewendet, um zu bestimmen, ob ein
10 Fetus mit cystischer Fibrose geboren werden wird, aber in der Vergangenheit wurde sie durchgeführt, nachdem ein bestimmtes Elternpaar bereits ein Kind mit cystischer Fibrose hatte, wodurch sie sicher als Träger identifiziert worden waren. In Kombination mit dem Trägernachweis, der vorstehend erläutert wurde, ist die DNA-Diagnose bei allen Schwangerschaften von Trägerpaaren möglich. Wenn die Eltern bereits ein Kind
15 mit cystischer Fibrose bekommen haben, kann eine extendierte Haplotypanalyse des Fetus durchgeführt werden, und somit wird der Prozentanteil an falschen positiven oder falschen negativen Ergebnissen stark verringert. Wenn die Eltern bisher kein erkranktes Kind bekommen haben und die DNA-Diagnose des Fetus aufgrund eines Trägernachweises durchgeführt wird, kann trotzdem eine Haplotypanalyse durchgeführt werden.

20

Obwohl man jahrelang der Meinung war, daß eine große klinische Heterogenität der Krankheit cystische Fibrose besteht, stellt sich nun heraus, daß es zwei Hauptkategorien gibt, die als normale Bauchspeicheldrüsenfunktion (CF-PS) und Bauchspeicheldrüsen-Unterfunktion (CF-PI) bezeichnet werden. Wenn die Mutationen, die mit
25 diesen Krankheitskategorien zusammenhängen, gut beschrieben sind, kann man eine bestimmte Mutation mit einem klinischen Phänotyp dieser Krankheit assoziieren. Dies ermöglicht Veränderungen der Behandlung jedes Patienten. Somit kann anhand der Natur der Mutation bis zu einem gewissen Grad eine Prognose für den Patienten vorausgesagt werden und eine spezifische Behandlung angegeben werden.

30



6.0 MOLEKULARBIOLOGIE DER CYSTISCHEN FIBROSE

Die Behauptung, daß CFTR die Aktivität der Ionenkanäle, insbesondere die des nach außen gerichteten C_1 -Kanals reguliert, der als funktioneller Defekt in CF impliziert wird, kann durch Injektion und Translation von in voller Länge *in vitro* transkribierter CFTR-mRNA in *Xenopus* Oocyten überprüft werden. Die sich daraus ergebenden Veränderungen der Ionenströme durch die Oocytusmembran können gemessen werden, wenn die Spannung bei einem bestimmten Wert gehalten wird. Es ist möglich, daß CFTR die endogenen Oocytumkanäle reguliert oder daß sie notwendig ist, um auch Epithelzell-RNA einzuführen, um die Translation der Kanalproteine zu leiten. Die Verwendung von mRNA, die für normale und für mutierte CFTR kodiert, wie in dieser Erfindung bereitgestellt, macht diese Versuche möglich.

Andere Arten der Expression in einem heterologen Zellsystem erleichtern ebenfalls die Aufgliederung von Struktur/Funktions-Beziehungen. Die vollständige CFTR-DNA-Sequenz, die in einen Plasmid-Expressionsvektor ligiert wurde, wird verwendet, um Zellen zu transfizieren, damit deren Einfluß auf den Ionentransport bestimmt werden kann. Plasmid-Expressionsvektoren, die den Teil einer normalen CFTR-Sequenz zusammen mit Teilen einer modifizierten Sequenz an ausgewählten Stellen enthalten, können bei *in vitro* Mutageneseversuchen verwendet werden, die durchgeführt werden, um diese Abschnitte des CFTR-Proteins zu identifizieren, die für die Regulationsfunktion von grundlegender Bedeutung sind.

6.1 EXPRESSION DER MUTIERTEN DNA-SEQUENZ

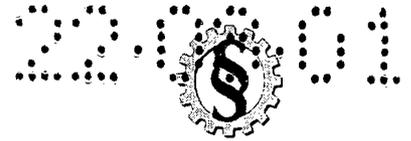
Die mutierte DNA-Sequenz kann in Untersuchungen manipuliert werden, um die Expression des Gens und seines Produkts zu verstehen und um die Produktion von großen Mengen des Proteins für die funktionelle Analyse, die Antikörperproduktion und die Patiententherapie zu erreichen. Die Veränderungen in der Sequenz können, aber müssen nicht, die Expressionsmuster bezüglich der relativen Mengen, der Gewebespezifität und der funktionellen Eigenschaften ändern. Die Teil- oder ungekürzten



cDNA-Sequenzen, die für das Protein des Subjekts kodieren, können, unmodifiziert oder modifiziert, an bakterielle Expressionsvektoren, wie pRIT- (Nilsson et al., *EMBO J.* 4: 1075-1080 (1985)), pGEX- (Smith und Johnson, *Gene* 67: f31-40 81988)) oder pATH- (Spindler et al., *J. Virol.* 49: 132-141 (1984)) Plasmide ligieren, die in *E. coli*-
5 Zellen für die Produktion der entsprechenden Proteine eingeführt werden können, die auch gemäß den zuvor erörterten Proteinreinigungsverfahren isoliert werden können. Die DNA-Sequenz kann auch aus ihrem existierenden Kontext zu anderen Klonierungsträgern, wie anderen Plasmiden, Bakteriophagen, Cosmiden, tierischen Viren, künstlichen Hefechromosomen (YAC) (Burke et al., *Science* 236: 806-812, (1987)), Soma-
10 zellen und anderen einfachen oder komplexen Organismen, wie Bakterien, Pilzen (Timberlake und Marshall, *Science* 244: 1313-1317 (1989)), Wirbellosen, Pflanzen (Gasser und Fraley, *Science* 244: 1293 (1989)), und Schweinen (Pursel et al., *Science* 244: 1281-1288 (1989)) transferiert werden.

15 Für die Expression in Säugerzellen kann die cDNA-Sequenz zu heterologen Promotoren ligiert werden, wie dem Affen-Virus (SV) 40, dem Promoter im pSV2-Vektor [Mulligan und Berg, *Proc. Natl. Acad. Sci. USA*, 78: f2072-2076 (1981)] und in Zellen eingeführt werden, wie Affen-COS-1-Zellen [Gluzman, *Cell* 23: 175-182 (1981)], um eine vorübergehende oder eine Langzeit-Expression zu erhalten. Die stabile
20 Integration des Chimären-Genkonstrukts kann in Säugerzellen durch biochemische Selektion, wie Neomycin [Souther and Berg, *J. Mol. Appln. Genet.* 1:327-341 (1982)] und Mycophenolsäure [Mulligan und Berg, s.o.] aufrecht erhalten werden.

Die DNA-Sequenzen können mit Standardverfahren manipuliert werden, wie der
25 Restriktionsenzym-Digestion, dem Auffüllen mit DNA-Polymerase, der Deletion durch Exonuclease, der Extension durch terminale Desoxynucleotid-Transferase, der Ligation von synthetischen oder klonierten DNA-Sequenzen, der ortsspezifischen Sequenzveränderung durch einsträngige Bakteriophagen-Zwischenstufen oder der Verwendung von spezifischen Oligonucleotiden in Kombination mit PCR.



Die cDNA-Sequenz (oder Abschnitte, die von ihr erhalten wurden) oder ein Mini-
gen (eine cDNA mit einem Intron und seinem eigenen Promotor) wird in eukaryon-
tische Expressionsvektoren durch herkömmliche Verfahren eingeführt. Diese Vektoren
sind so ausgelegt, daß sie die Transkription der cDNA in eukaryontischen Zellen ermög-
5 lichen, indem sie regulatorische Sequenzen bereitstellen, die die Transkription der
cDNA initiieren und verstärken, und ihre angemessene Aufspaltung und Polyadenylierung
sicherstellen. Vektoren, die die Promoter- und Verstärkerregionen des Affenvirus
(SV) 40 oder eine lange terminale Wiederholung (LTR) des Rous Sarcoma Virus und
ein Polyadenylations- und Aufspaltungssignal von SV 40 enthalten, sind ohne weiteres
10 erhältlich [Mulligan et al., *Proc. Natl. Acad. Sci. USA* 78: 1078-2076, (1981); Gorman
et al., *Proc. Natl. Acad. Sci. USA* 79: 6777-6781 (1982)]. Andererseits kann auch der
endogene CFTR-Promotor verwendet werden. Der Expressionsgrad der cDNA kann mit
diesem Vektortyp manipuliert werden, entweder indem man Promotoren verwendet, die
unterschiedliche Aktivitäten aufweisen (beispielsweise kann der Baculovirus pAC373
15 cDNAs in großem Umfang in *S. Fungiperda*-Zellen exprimieren [M. D. Summers und
G.E. Smith in *Genetically Altered Viruses and the Environment* (B. Fields et al.,
Herausg.) Bd. 22, Nr. 319-328, Cold Spring Harbour, New York, 1985] oder durch Ver-
wendung von Vektoren, die Promotoren aufweisen, die für eine Modulation empfäng-
lich sind, beispielsweise den Glucocorticoid-verantwortlichen Promotor aus dem
20 Mäuse-Mammatumovirus [Lee et al., *Nature* 294:228 (1982)]. Die Expression der
cDNA kann in den Rezipientenzellen 24 bis 72 Stunden nach der Einführung beob-
achtete werden (transiente Expression) .

Darüber hinaus enthalten einige Vektoren selektierbare Marker [wie die gpt-
25 [Mulligan et Berg, s.o.] oder neo- [Southern und Berg, *J. Mol. Appln. Genet.* 1: 327-341
(1982)] Bakteriengene, die die Isolierung von Zellen, welche stabile Langzeitexpressio-
nen der Vektoren (und damit der cDNA) in der Rezipientenzelle aufweisen, durch
chemische Selektion ermöglichen. Die Vektoren können in den Zellen als episomale,
frei replizierende Einheiten beibehalten werden, indem man regulatorische Elemente
30 von Viren verwendet, wie Papilloma[Sarver et al., *Mol. Cell. Biol.* 1:486 (1981)] oder
Epstein-Barr (Sugden et al., *Mol. Cell Biol.* 5:410 (1985)]. Andererseits kann man auch



Zelllinien erzeugen, bei denen der Vektor in die genomische DNA integriert ist. Beide dieser Zelliniertypen erzeugen das Genprodukt auf kontinuierlicher Basis. Man kann auch Zelllinien erzeugen, bei denen die Zahl der Kopien des Vektors (und damit auch die der cDNA) amplifiziert ist, um Zelllinien zu erzeugen, die das Genprodukt in großem
5 Umfang produzieren können [Alt et al., *J. Biol. Chem.* 253: 1357 (1978)].

Der Transfer von DNA in eukaryontische, insbesondere menschliche oder andere Säugerzellen, stellt derzeit ein herkömmliches Verfahren dar. Die Vektoren werden in die Rezipientenzellen als reine DNA (Transfektion) durch beispielsweise Ausfällung
10 mit Calciumphosphat [Graham und Vander Eb, *Virology* 52:466 81973] oder Strontiumphosphat [Brash et al., *Mol. Cell Biol.* 7:2013 (1987)], Elektroporation [Neumann et al., *EMBO J.* 1:841 (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)], Protoplastenfusion [Schafner,
15 *Proc. Natl. Acad. Sci. USA* 72:2163] oder Pelletpistolen [Klein et al., *Nature* 327: 70 (1987)]. Andererseits kann die cDNA durch Infektion mit Virusvektoren eingeführt werden. Es werden Systeme entwickelt, die beispielsweise Retroviren [Bernstein et al., *Genetic Engineering* 7: 235 (1985)9, Adenoviren [Ahmad et al., *J. Virol.* 57:267 (1986)] oder Herpesviren [Spate et al., *Cell* 30: 295 (1982)] verwenden.

20 Diese eukaryontischen Expressionssysteme können für viele Untersuchungen des mutierten CF-Gens und des mutierten CFTR-Produkts, beispielsweise an den Proteinpositionen 85, 148, 178, 455, 493, 507, 542, 549, 551, 560, 563, 574, 1077 und 1092, verwendet werden. Sie schließen beispielsweise ein: (1) den Nachweis, ob das Gen
25 angemessen exprimiert wird, und ob alle Posttranslations-Modifikationen, die für die vollständige biologische Aktivität notwendig sind, angemessen abgeschlossen wurden, (2) das Identifizieren von Regulationselementen, die in der 5'-Region des CF-Gens liegen, und ihre Rolle bei der Gewebe- oder temporalen Regulation der Expression des CF-Gens, (3) die Herstellung von großen Mengen des normalen Proteins zur Isolierung
30 und Reinigung, (4) die Verwendung von Zellen, die das CFTR-Protein exprimieren, als Assay-System für Antikörper, die für das CFTR-Protein erzeugt wurden, oder ein



Assay-System, für das Testen der Wirksamkeit von Arzneimitteln, (5) die Untersuchung der Funktion des normalen vollständigen Proteins, spezifischen Abschnitte des Proteins, oder von natürlich vorkommenden oder künstlich erzeugten mutierten Proteinen. Natürlich vorkommenden mutierte Proteine kommen in Patienten mit CF vor, während künstlich erzeugte mutierte Proteine durch ortsspezifische Sequenzänderungen konstruiert werden können. Diese letztgenannten Untersuchungen können die Funktion jedes beliebigen Aminosäurerests im Protein durch Mutieren des Nucleotids, das für diese Aminosäure kodiert, untersuchen.

10 Mittels der vorstehend genannten Verfahren können die Expressionsvektoren, die die mutierte CF-Gensequenz oder Fragmente davon enthalten, je nach Wunsch in menschliche Zellen, Säugerzellen anderer Spezies oder nicht-Säuger-Zellen eingeführt werden. Die Auswahl der Zelle wird vom Behandlungszweck bestimmt. Beispielsweise kann man Affen-COS-Zellen [Gluzman, *Cell* 23: 175 (1981)] verwenden, die große
15 Mengen des SV40 T-Antigens erzeugen und die Replikation von Vektoren, die den Replikationsursprung von SV40 enthalten, verwenden, um zu zeigen, daß der Vektor das Proteinprodukt exprimieren kann, da die Funktion nicht erforderlich ist. Eine ähnliche Behandlung könnte mit chinesischen Hamsterovarien (CHO) oder Mäuse NIH 3T3-Fibroblasten oder mit menschlichen Fibroblasten oder Lymphoblasten durchgeführt
20 werden.

Der rekombinante Klonierungsvektor, gemäß dieser Erfindung, umfaßt dann die selektierte DNA der DNA-Sequenzen dieser Erfindung zur Expression in einem geeigneten Wirt. Die DNA wird im Vektor operativ mit einer Expressions-Steuerungssequenz im rekombinanten DNA-Molekül verbunden, so daß normales CFTR-Polypeptid exprimiert werden kann. Die Expressions-Steuerungssequenz kann aus ausgewählt sein aus der Gruppe bestehend aus Sequenzen, die die Expression von Genen prokaryontischer oder eukaryontischer Zellen oder eukaryontischer Zellen und ihrer Viren und Kombinationen davon steuern. Die Expressions-Steuerungssequenz kann spezifisch
25 ausgewählt sein aus der Gruppe bestehend aus dem lac-System, dem trp-System, dem tyc-System, dem trc-System, Operator- und Promotor-Hauptregionen von Lambda-



Phagen, der Steuerungsregion von fd-Hüllprotein, den frühen und späten SV40-Pro-
motoren, die von Polyoma, Adenovirus, Retrovirus, Baculovirus und Affenvirus abge-
leitet wurden, dem Promotor für 3-Phosphoglyceratkinase, den Promotoren für Hefe-
säurephosphatase, dem Promotor für die Hefe-alpha-Kreuzungsfaktoren und Kombi-
5 nationen davon.

Die Wirtszelle, die mit dem Vektor dieser Erfindung transfektiert sein kann, kann
ausgewählt werden aus der Gruppe bestehend aus *E. coli*, *Pseudomonas*, *Bacillus sub-*
tillis, *Bacillus stearothermophilus* und anderen Bazillen; anderen Bakterien; Hefe;
10 Pilzen; Insekten; Mäusen und anderen Tieren; oder Pflanzenwirten; oder menschlichen
Gewebezellen.

Man ist sich dessen bewußt, daß für die mutierten DNA-Sequenzen ähnliche
Systeme angewendet werden, um das mutierte Produkt zu exprimieren und zu erzeugen.

15

6.2 ÜBERLEGUNGEN ZUR PROTEINFUNKTION

Um die Funktion des mutierten CFTR-Proteins zu untersuchen, verwendet man
vorzugsweise Epithelzellen als Rezipienten, da eine angemessene funktionale Ex-
20 pression die Anwesenheit von anderen Bahnen oder Genprodukten erforderlich machen
kann, die nur in solchen Zellen exprimiert werden. Zellen, die verwendet werden
können, sind unter anderem beispielsweise menschliche Epithelzelllinien, wie T84
(ATCC Nr. CRL 248) oder PANC-1 (ATCC Nr. CLL 1469) oder die immortalisierte
T43 CF-Nasalepithelzelllinie [Jettan et al., *Science* (1989)] und primäre [Yanhoskes et
25 al., *Ann. Rev. Resp. Dis.* 132: 1281 (1985)] oder transformierte [Scholte et al., *Exp. Cell.*
Res. 182: 559 (1989)] menschliche Nasenpolypen- oder Atemwegszellen, Pancreaszellen
[Harris und Coleman, *J. Cell. Sci.* 87: 695 (1987)] oder Schweißdrüsenzellen [Collie et
al., *In Vitro* 21:597 (1985)], die von normalen oder CF-Subjekten erhalten wurden. Die
CF-Zellen können verwendet werden, um die funktionelle Aktivität des mutierten CF-
30 Gens zu testen. Gegenwärtig zur Verfügung stehende funktionelle Assays beinhalten die
Untersuchung der Bewegung der Anionen (Cl oder I) durch die Zellmembrane als



Funktion der Stimulation der Zellen durch Agentien, die die intrazellulären AMP-Niveaus erhöhen und die Chloridkanäle aktivieren [Stutto et al., *Proc. Nat. Acad. Sci. U.S.A.*, 82: 6677 (1985)]. Andere Assays beinhalten die Messung von Veränderungen des Zellpotentials durch Patch-Clamping von ganzen Zellen oder von isolierten Membranen [Frizzell et al., *Science* 233: 558 (1986), Welsch und Liedtke, *Nature* 322: 467 (1986)] oder die Untersuchung von Ionenströmen in Epithelschichten von konfluenten Zellen [Widdicombe et al., *Proc. Nat. Acad. Sci.* 82: 6167 (1985)]. Andererseits könnte RNA, die aus dem CF-Gen hergestellt wurde, in *Xenopus* Oocyten eingespritzt werden. Der Oocyt translatiert RNA in Protein und ermöglicht deren Untersuchung. DA andere spezifischere Assays entwickelt werden, können sie bei der Untersuchung der Funktion des transfektierten mutierten CFTR-Proteins ebenfalls verwendet werden.

“Domain-Umschalt“-Versuche zwischen mutierter CFTR und dem menschlichen Arzneimittel-Mehrfachresistenz-P-Glycoprotein können ebenfalls durchgeführt werden, um die Untersuchung des mutierten CFTR-Proteins weiterzuführen. In diesen Versuchen werden Plasmid-Expressionsvektoren durch Routineverfahren aus Fragmenten der mutierten CFTR-Sequenz und Fragmenten der Sequenz von P-Glycoprotein konstruiert, die durch DNA-Ligase verbunden wurden, so daß ein Protein, das die jeweiligen Abschnitte dieser beiden Proteine enthält, durch eine Wirtszelle, die mit dem Plasmid transfiziert wurde, synthetisiert wird. Der letztgenannte Ansatz hat den Vorteil, daß viele experimentelle Parameter, die mit der Arzneimittel-Mehrfachresistenz assoziiert sind, gemessen werden können. Somit ist es nunmehr möglich, die Fähigkeit von Segmenten von mutierter CFTR, diese Parameter zu beeinflussen, zu bestimmen.

Diese Untersuchungen des Einflusses von mutierter CFTR auf den Ionentransport dient dazu, das Gebiet des Epitheltransports in die molekulare Diskussion einzubringen.

6.3 THERAPIEN

Selbstverständlich ist die Entwicklung von Therapien, um den CF-Defekt zu umgehen oder zu überwinden, das Hauptziel der verschiedenen biochemischen Unter-



suchungen der Zusammensetzungen dieser Erfindung, wobei sowohl pharmakologische als auch "Gentherapie"-Methoden angewendet werden

Bei der pharmakologischen Methode wird nach Arzneimitteln, die den CF-Defekt
5 umgehen oder überwinden, gesucht. Anfangs können Verbindungen im wesentlichen
nach dem Zufallsprinzip getestet werden, und Screening-Systeme sind erforderlich, um
zwischen vielen in Frage kommenden Verbindungen zu unterscheiden. Diese Erfindung
stellt Wirtszellsysteme bereit, die viele der mutierten CF-Gene exprimieren, die sich
insbesondere gut für die Verwendung als Screening-Systeme der ersten Stufe eignen.
10 Vorzugsweise wird ein Zellkultursystem, bei dem Säugerzellen (am bevorzugtesten
menschliche Zellen), die mit einem Expressionsvektor transfektiert wurden, der eine
DNA-Sequenz enthält, die für das CFTR-Protein kodiert, das eine CF-bewirkende
Mutation, beispielsweise die I507-Deletion, enthält, in dem Screening-Verfahren ver-
wendet. In Frage kommende Arzneimittel werden getestet, indem an Zellen in An-
15 wesenheit des in Frage kommenden Arzneimittels inkubiert und die Zellfunktionen
mißt, die von CFTR abhängen, insbesondere durch Messen der Ionenströme, wo das
Transmembranpotential bei einem bestimmten Werte gehalten wird. Um mit der großen
Zahl von Assay zurechtzukommen, basieren bequemere Assays jedoch beispielsweise
auf der Verwendung von Ionen-empfindlichen fluoreszierenden Farblösungen. Um
20 Veränderungen in der Cl-Ionenkonzentration nachzuweisen, sind SPQ oder dessen Ana-
loge nützlich.

Alternativ dazu könnte ein zellfreies System verwendet werden. Gereinigte CFTR
könnte in künstlichen Membranen wiederhergestellt werden und die Arzneimittel
25 könnten in einem zellfreien Assay gescreent werden [Al-Aqwatt, *Science* (1989)].

In der zweiten Stufe ist der Tierversuch notwendig. Es ist möglich, ein CF-Modell
zu entwickeln, indem man in die normale Expression des Gegenstücks zum CF-Gen in
einem Tier, wie einer Maus, eingreift. Der "Knock out" dieses Gens durch Einführung
30 einer mutierten Form davon in die Keimbahn von Tieren liefert einen Tierstamm mit



CF-ähnlichen Symptomen. Dies ermöglicht das Testen von Arzneimitteln, die sich auf der ersten, auf Zellen basierenden Stufe als vielversprechend erwiesen haben.

Da weitere Erkenntnisse über die Natur des Proteins und seine Funktion erhalten werden, wird es möglich sein, die Strukturen von Proteinen oder anderen Verbindungen vorauszusagen, die mit dem CFTR-Protein interagieren. Dies wird wiederum bestimmte Voraussagen über potentiellen Arzneimittel ermöglichen, die mit diesem Protein interagieren, und sich auf die Behandlung der Patienten auswirken. Schließlich können solche Arzneimittel auf der Grundlage von vorausgesagten Strukturen, die nötig sind, um mit CFTR-Domänen zu interagieren, entwickelt und chemisch synthetisiert werden. Dieser Ansatz ist in Capsey und Delvatte, *Genetic Engineered Human Therapeutic Drugs*, Stockton Press, New York, 1988 besprochen. Diese potentiellen Arzneimittel müssen ebenfalls im Screening-System getestet werden.

6.3.1 PROTEINAUSTAUSCH-THERAPIE

Die Behandlung von CF kann durch Ersetzen des defekten Proteins mit normalem Protein, durch Modulieren der Funktion des defekten Proteins oder durch Modifizieren einer anderen Stufe des Stoffwechselwegs, an dem CFTR beteiligt ist, um die physiologische Abnormalität zu korrigieren, durchgeführt werden.

Um in der Lage zu sein, das defekte Protein durch die normale Version zu ersetzen, muß man ausreichend große Mengen des reinen CFTR-Proteins besitzen. Reines Protein kann wie an früherer Stelle beschrieben aus kultivierten Zellsystemen erhalten werden. Die Lieferung des Proteins zum befallenden Atemwegsgewebe erfordert dessen Packung in Lipid-haltige Träger, die den Einbau des Proteins in die Zellmembran erleichtern. Es kann auch möglich sein, Träger zu verwenden, die Proteine, wie Oberflächenprotein einbauen, beispielsweise SAP (Val) oder SAP (phe), die diese Funktion von Natur aus erfüllen, zumindest für alveolare Lungenzellen. (PCT-Patentanmeldung WO/8803170, Whitsett et al., 7. Mai 1988, und PCT-Patentanmeldung WO89/04327, Benson et al., 18. Mai 1989). Die CFTR-haltigen Träger werden in die Atemwege durch



Inhalation oder Irrigation eingeführt, Techniken, die gegenwärtig bei der CF-Behandlung angewendet werden (Boat et al., s.o.).

6.3.2 ARZNEIMITTEL THERAPIE

5

Die Modulation der CFTR-Funktion kann durch die Verwendung von Therapeutika (Arzneimitteln) durchgeführt werden. Diese können nach dem Zufallsprinzip mittels eines Screening-Programms, bei dem ihre Wirksamkeit bei der Modulation des defekten CFTR-Proteins *in vitro* beobachtet wird, identifiziert werden. Screening-Programme können kultivierte Zellsysteme verwenden, in denen das defekte CFTR-Protein exprimiert wird. Als Alternative können Arzneimittel entwickelt werden, die die CFTR-Aktivität anhand des Wissens um die Struktur- und Funktions-Korrelationen des CFTR-Proteins und anhand des Wissens über den spezifischen Defekt im mutierten CFTR-Protein mutieren (Capsey und Delvatte, s.o.). Es ist möglich, daß das mutierte CFTR-Protein ein anderes Arzneimittel für eine bestimmte Modulation erfordert. Dann ist es notwendig, die spezifische(n) Mutation(en) in jedem CF-Patienten vor dem Beginn der Arzneimitteltherapie zu identifizieren.

Es können Arzneimittel entwickelt werden, die auf verschiedene Aspekte der CFTR-Proteinstruktur oder -funktion wirken. Beispielsweise kann ein Arzneimittel (oder ein Antikörper) an eine strukturelle Falte des Proteins binden, um eine defekte Struktur zu korrigieren. Alternativ dazu könnte ein Arzneimittel an einen spezifischen funktionellen Rest binden und dessen Affinität für ein Substrat oder eine Cofaktor erhöhen. Da es bekannt ist, daß Mitglieder der Proteinklasse, für die CFTR eine strukturelle Homologie besitzt, mit einer Reihe von Arzneimitteln interagieren, diese binden und transportieren können, kann man vernünftigerweise annehmen, daß auf Arzneimitteln beruhende Therapien bei der Behandlung von CF wirksam sind.

Ein dritter Mechanismus zur Verstärkung der Aktivität eines wirksamen Arzneimittels bestünde darin, die Produktion oder die Stabilität von CFTR im Inneren der



Zelle zu modulieren. Diese Erhöhung der CFTR-Menge könnte dessen defekte Funktion kompensieren.

Die Arzneimitteltherapie kann auch angewendet werden, um die defekte CFTR-
5 Funktion durch Wechselwirkung mit anderen Komponenten des physiologischen oder
biochemischen Stoffwechselwegs, der für die Expression der CFTR-Funktion not-
wendig ist, zu kompensieren. Diese Wechselwirkungen können zu einer Steigerung oder
einer Abnahme der Aktivität in diesen Hilfsproteinen führen. Die Verfahren für die
Identifikation dieser Arzneimittel wären denen ähnlich, die vorstehend für Arzneimittel
10 im Zusammenhang mit CFTR beschrieben wurden.

Bei anderen genetischen Störungen war es möglich, die Folgen geänderter oder
fehlender normaler Funktionen durch Anpassung der Ernährung zu korrigieren. Dies
fand statt, indem man Metaboliten eliminierte, wie im Fall von Phenylketonuria, wo
15 Phenylalanin bei der Diät in den ersten fünf Lebensjahren weggelassen wird, um eine
geistige Entwicklungsverzögerung zu verhindern, oder durch die Zugabe großer
Mengen von Metaboliten zur Diät, wie im Fall von Adenosin-Desaminase, wo die
funktionelle Korrektur der Enzymaktivität durch die Zugabe des Enzyms zur Diät
korrigiert werden kann. Somit kann, sobald die Einzelheiten der CFTR-Funktion
20 verstanden sind und der Grunddefekt in CF definiert sein wird, eine Therapie mittels
diätetischer Manipulationen möglicherweise erreicht werden.

Der zweite potentielle therapeutische Ansatz besteht in der sogenannten "Gen-
therapie", bei der normale Kopien des CF-Gens in Patienten eingeführt werden, um
25 erfolgreich für normales Protein in den wichtigsten Epithelzellen der befallenen Gewebe
zu kodieren. Es ist äußerst wichtig, dies bei den Epithelzellen der Atemwege des oberen
Atemungstrakts zu versuchen. Das CF-Gen wird diesen Zellen in einer Form zugeführt,
in der es aufgenommen werden kann und für genügend Protein kodieren kann um eine
Regulationsfunktion bereitzustellen. Als Ergebnis davon können die Lebensqualität und
30 die Lebensdauer des Patienten stark verlängert werden. Schließlich besteht das Ziel
selbstverständlich darin, das Gen allen befallenen Geweben zuzuführen.



6.3.3 GENTHERAPIE

Ein Ansatz zur Therapie von CF besteht in der Insertion von einer normalen
5 Version des CF-Gens in das Atemwegsepithelium erkrankter Patienten. Es ist von Be-
deutung, daß das obere Atemwegssystem die Hauptursache für die Morbidität und
Mortilität bei CF darstellt; obwohl die Pancreas-Erkrankung ein Hauptmerkmal ist,
kann sie heutzutage durch Enzymsupplementation relativ gut behandelt werden. Somit
würde eine Körperzellentherapie [eine Besprechung findet sich in T. Friedmann,
10 *Science* 244: 1275 (1989)], die auf die Atemwege gerichtet ist, die schwerwiegendsten
Probleme im Zusammenhang mit CF lindern.

A. RETROVIRUSVEKTOREN. Retroviren wurden als bevorzugte Vektoren für
Versuche mit der Körperzellentherapie mit einer hohen Infektionseffizienz und stabiler
15 Integration und Expression angesehen [Orkin et al., *Proc. Med. Genet.* 7:130 (1988)].
Ein möglicher Rückschlag besteht darin, daß eine Zellteilung für die Retrovirusintegra-
tion notwendig ist, so daß die Zielzellen in den Atemwegen vor der Retrovirusinfektion
möglicherweise in den Zellzyklus geschoben werden müssen, möglicherweise auf
chemischem Wege. Die CF-Gen-DNA kann in voller Länge in einen Retrovirusvektor
20 kloniert werden und entweder von ihrem endogenen Promotor oder von der retroviralen
LRT (langen terminalen Wiederholung) getrieben werden. Die Expression von
normalem Protein in einer Menge von 10 % des endogenen mutierten Proteins würde
man als günstig ansehen, da es sich hier um eine rezessive Krankheit handelt. Die
Zufuhr des Virus könnte durch Aerosol oder Instillation in die Trachea durchgeführt
25 werden.

B. ANDERE VIRUSVEKTOREN Andere Zufuhrsysteme, die verwendet wer-
den können, sind unter anderem der Adeno-assoziierte Virus [ASV, McLaughlin et al.,
J. Virol. 62: 1963 (1988)], der Vacciniavirus [Moss et al., *Annu. Rev. Immunol.* 5: 305,
30 1987)], der Rinder-Papillomavirus [Rasmussen et al, *Method Enzymol.* 139: 642 (1987)]
oder Mitglieder der Herpesvirusgruppe, wie der Epstein-Barr-Virus (Margolskee et al.,



Mol. Cell. Biol. 8: 2937 (1988)]. Obwohl noch viel über ihre grundlegende Biologie zu lernen wäre, ist die Idee der Verwendung eines Virusvektors mit natürlichem Topismus für den oberen Atemwegstrakt (z.B. der synzytiale Atemwegsvirus, der Echovirus, des Coxsackie-Virus etc.) möglich.

5

C. NICHT-VIRALER GENTRANSFER Es können auch andere Methoden zur Insertion des CF-Gens in das Epithelium der oberen Atemwege zum Erfolg führen; viele davon weisen eine geringere Effizienz auf und würden möglicherweise eine Infektion *in vitro*, eine Selektion von Transfektanten und eine Reimplantation erfordern.

10 Dies würde Calciumphosphat, DEAE-Dextran, die Elektroporation und die Protoplastenfusion einschließen. Eine besonders vielversprechende Idee ist die Verwendung von Liposom, die möglicherweise *in vivo* durchgeführt werden könnte [Ostro, *Liposomes*, Marcel-Dekker, 1987]. Synthetische kationische Lipide, wie DOTMA [Felger et al., *Proc. Natl. Acad. Sci, USA* 84:7413 (1987)] könnten die Wirksamkeit und die
15 Leichtigkeit der Durchführung dieser Methode verbessern.

6.4 CF-TIERMODELLE

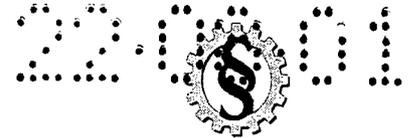
Die Erschaffung eines Mäuse- oder eines anderen Tiermodells für CF ist von
20 größter Wichtigkeit für das Verständnis der Krankheit und für das Testen möglicher Therapien (ein allgemeiner Überblick über die Bereitstellung von Tiermodellen findet sich in Erickson, *Am. J. Hum. Genet.* 43:582 (1988)]. Gegenwärtig existiert kein Tiermodell für CF. Die evolutionäre Konservierung des CF-Gens (wie durch die Spezieskreuzungs-Hybridisationsblots für E4.3 und H1.6 gezeigt wurde), wie in Fig. 4 gezeigt,
25 weist darauf hin, daß ein orthologes Gen in der Maus vorhanden ist (im folgenden als mCF bezeichnet, sein entsprechendes Protein als mCFTR bezeichnet), und es wäre möglich dieses in genomischen und cDNA-Mäusebibliotheken zu klonieren, wobei die menschlichen CF-Gensonden verwendet werden. Man erwartet, daß die Generation einer spezifischen Mutation im Mäusegen analog zur I507-Mutation am geeignetsten
30 wäre, um den Phänotyp zu reproduzieren, obwohl die Erzeugung einer vollständigen Inaktivierung des mCFTR-Gens ebenfalls ein nützlicher Mutant wäre.



A. MUTAGENESE Die Inaktivierung des mCF-Gens kann durch chemische [z.B. Johnson et al., *Proc. Natl. Acad. Sci. USA* 78:3138 (1981)] oder Röntgenstrahlen-Mutagenese [Popp et al., *J. Mol. Biol.* 127:141 (1979)] von Mäusegameten, gefolgt von Fertilisation, durchgeführt werden. Heterozygoter Nachwuchs für die Inaktivierung von mCFTR kann dann durch Southern Blotting identifiziert werden, um den Verlust eines Allels durch Dosierung, oder das Nichtvererben eines elterlichen Allels, wenn ein RFLP-Marker festgestellt wird, zu zeigen. Dieser Ansatz wurde kürzlich mit Erfolg angewendet, um Mäusemutanten für α -Globin zu identifizieren [Whitney et al., *Proc. Natl. Acad. Sci. USA* 77: 1087 (1980)], Phenylalaninhydroxylase [McDonald et al., *Pediatr. Res.* 23:63 (1988)] und Carbonatdehydrase II [Lewis et al., *Proc. Natl. Acad. Sci. USA*, 85:1962, (1988)].

B. TRANSGENE Eine mutierte Version von CFTR oder Mäuse-CFTR kann in die Mäusekeimbahn inseriert werden unter Verwendung derzeitiger Standardtechniken der Oocyteninjektionen [Camer, *Trends in Genetics* (1988)]; alternativ dazu kann, falls es erwünscht ist, das endogene mCF-Gen zu inaktivieren oder zu ersetzen, das homologe Rekombinationssystem, das Embryonen (ES) -Stammzellen verwendet [Capecchi, *Science* 244: 1288 81989)], angewandt werden.

1. OOCYTENINJEKTION Das Plazieren einer oder mehrerer Kopien des normalen oder mutierten mCF-Gens an einer zufälligen Stelle in der Mäusekeimbahn kann durch Mikroinjektion des Pronucleus eines gerade befruchteten Mäuseoocyten, gefolgt von Reimplantation in eine scheinträchtige Leihmutter durchgeführt werden. Die lebendgeborenen Mäuse können dann auf der Suche nach integrierten Elementen mittels Analyse der Schwanz-DNA auf Anwesenheit menschlicher CF-Gensequenzen hin gescreent werden. Das gleiche Protokoll kann verwendet werden, um ein mutiertes mCF-Gen zu inserieren. Um ein Mäusemodell zu erschaffen, würde man dieses Transgen in einen Mäusehintergrund plazieren wollen, wo das endogene mCF-Gen inaktiviert wurde, entweder durch Mutagenese (siehe oben) oder durch homologe Rekombination (siehe unten). Bei dem Transgen kann es sich um eines der folgenden handeln: a) eine voll-



ständige genomische Sequenz, obwohl deren Größe (etwa 250 kb) erforderlich machen würde, daß sie als künstliches Hefechromosom oder als Chromosomenfragment injiziert würde; b) eine cDNA mit entweder dem natürlichen Promotor oder einem heterologen Promotor; c) ein "Minigen", das die ganze Kodierungsregion und verschieden andere Elemente, wie Introns, Promotor und 3'-flankierende Elemente umfaßt, die sich als notwendig für die optimale Expression erwiesen haben.

2. RETROVIRUSINFEKTION VON FRÜHEN EMBRYOS

Diese Alternative

beinhaltet das Insertieren des CFTR- oder mCF-Gens in einen Retrovirusvektor und das direkte Infizieren von Mäuseembryos in frühen Entwicklungsstadien, wodurch eine Chimäre erschaffen wird [Soriano et al., *Cell* 46:19 (1986)]. Zumindest einige davon führen zu einer Keimbahn-Transmission.

3. ES-ZELLEN UND HOMOLOGE REKOMBINATION

Der Embryonen-

stammzellen-Ansatz (Capecchi, s.o., und Capecchi, *Trends Genet.* 5:70 (1989)) eröffnet die Möglichkeit zur Durchführung eines Gentransfers und das anschließende Screening der resultierenden totipotenten Zellen, um die seltenen homologen Rekombinationseignisse zu identifizieren. Sobald sie identifiziert sind, können sie verwendet werden, um Chimären zu erzeugen durch Injektion von Mäuseblastocyten und ein Teil der resultierenden Mäuse wird eine Keimbahntransmission von der rekombinanten Bahn zeigen. Es gibt mehrere Arten, wie man sich dies bei der Erzeugung eines Mäusemodells für CF zunutze machen könnte:

- a) Die Inaktivierung des mCF-Gens kann auf bequeme Weise durchgeführt werden, indem man ein DNA-Fragment entwickelt, das Sequenzen aus einem mCFTR-Exon, das einen selektierbaren Marker flankiert, wie neo, enthält. Homologe Rekombination wird zur Insertion der neo-Sequenzen in der Mitte eines Exons führen, wodurch mCFTR inaktiviert wird. Die homologe Rekombination (üblicherweise etwa 1 von 1000) kann von den heterologen durch DNA-Analyse von individuellen Klonen unterschieden werden [üblicherweise mittels PCR, Kim et al. *Nucleic Acids Res.* 16: 8887 (1988), Joyner et al., *Nature* 338: 153 81989);



Zimmer et al., s.o., S. 15) oder durch Anwendung einer negativen Selektion gegenüber den heterologen Ereignissen [so wie die Verwendung eines HSV TK-Gens am Ende des Konstrukts, gefolgt von der Gancyclovir-Selektion, Nasour et al., *Nature* 336: 348 (1988)]. Diese inaktivierte mCFTR-Maus kann dann verwendet werden, um ein mutiertes CF-Gen oder mCF-Gen, das beispielsweise die I507-Abnormität oder irgend eine andere erwünschte Mutation trägt, einzuführen.

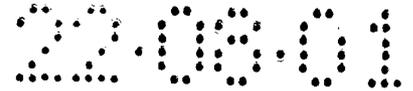
b) Es ist möglich, daß spezifische Mutanten von mCFTR-cDNA in einem Schritt erzeugt werden. Beispielsweise kann man ein Konstrukt herstellen, das mCF-Intron 9 Sequenzen am 5'-Ende, ein selektierbares neo-Gen in der Mitte und Intron 9 + Exon 10 (das die Mäuseversion der I507-Mutation enthält) am 3'-Ende enthält. Ein homologes Rekombinationsereignis würde zur Insertion des neo-Gens in Intron 9 und die Ersetzung von Exon 10 durch die mutierte Version führen.

c) In Anwesenheit des selektierbaren neo-Markers in der geänderten Intron-expression des mCF-Gens wäre es möglich, es in einem zweiten homologen Rekombinationsschritt zu excisieren.

d) Es ist auch möglich, Mutationen in der Mäusekeimbahn zu erzeugen, indem man Oligonucleotide, die die Mutation von Interesse enthalten, injiziert und die resultierenden Zellen durch PCR screent.

Diese Ausführungsform der Erfindung bezog sich in erster Linie auf ein Mäusemodell der cystischen Fibrose. Fig. 4 zeigt Spezieskreuzungs-Hybridisation nicht nur an Mäuse-DNA, sondern auch an Rinder-, Hamster- und Hühner-DNA. Somit denkt man darüber nach, daß ein orthologes Gen auch in vielen anderen Spezies existiert. Es wird daher erwogen, daß es möglich sein wird, andere Tiermodelle unter Verwendung ähnlicher Techniken zu erzeugen.

30



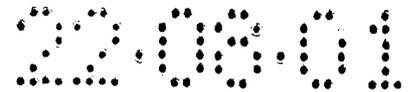
Ansprüche

1. DNA-Molekül, umfassend eine intronfreie DNA-Sequenz, die für ein mutiertes
5 CFTR-Polypeptid codiert, mit der Sequenz gemäß Fig. 4 für die Aminosäurerest-
Positionen 1 bis 1480 und das außerdem durch Nucleotidsequenzvarianten
gekennzeichnet ist, die zu einer Deletion oder Änderung von Aminosäuren an min-
destens einer Restposition, ausgewählt aus der Gruppe bestehend aus 85, 148, 178,
455, 493, 507, 542, 549, 560, 563, 574, 1077 und 1092 führen, wobei es sich bei
10 einer Änderung an der Position 549 entweder um S549R oder S549I handelt.

2. DNA-Molekül, umfassend eine DNA-Sequenz, die für ein mutiertes CFTR-Poly-
peptid kodiert, mit der Sequenz gemäß Fig. 1 für die DNA-Sequenzpositionen 1
bis 4575, und das außerdem durch Nucleotidsequenzvarianten gekennzeichnet ist,
15 die zu einer Deletion oder Änderung der DNA an mindestens einer DNA-Sequenz-
position, ausgewählt aus der Gruppe bestehend aus 129, 556 und 3659 der Fig. 1
oder einer Deletion oder Änderung der Sequenz der Introns 4, 5 und 10, wie in
Fig. 18 gezeigt, an den Nucleotidpositionen, die von 621+1, 711+1 und 1717-1
der Fig. 1 dargestellt sind, führen.

- 20 3. DNA-Molekül, umfassend eine DNA-Sequenz, ausgewählt aus der Gruppe be-
stehend aus:
 - (a) DNA-Sequenzen, die einem Fragment einer selektierten Sequenz nach An-
spruch 1 oder 2 entsprechen, einschließlich von mindestens 16 Nucleotiden;
25 und
 - (b) DNA-Sequenzen, die für ein Epitop kodieren, dadurch gekennzeichnet, daß
das mutierte CFTR-Protein von mindestens 18 sequentiellen Nucleotiden
kodiert wird, wobei zumindest eine der deletierten oder geänderten Positio-
nen in der selektierten Sequenz nach Anspruch 1 oder 2 liegt.

30

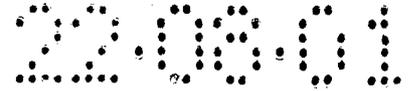


4. Nucleinsäuresonde, umfassend eine markierte Nucleotidsequenz, die der selektierten Sequenz entspricht, die in den Teilen (a) oder (b) von Anspruch 3 genannt ist.
5. Rekombinations-Klonierungsvektor, umfassend das DNA-Molekül nach Anspruch 3.
6. Nicht-menschlicher Wirt, transformiert mit dem Vektor gemäß Anspruch 5.
7. Verfahren zum Erzeugen eines mutierten CFTR-Polypeptids, umfassend die folgenden Stufen:
 - (a) Kultivieren einer Wirtszelle, die vom Vektor nach Anspruch 5 transfektiert wurde, in dem Medium und unter Bedingungen, die die für die Expression des mutierten CFTR-Polypeptids günstig sind;
 - (b) Isolieren des exprimierten mutierten CFTR-Polypeptids.
8. Gereinigtes mutiertes CFTR-Polypeptid, das von der in Anspruch 3 genannten DNA-Sequenz kodiert wird.
9. Verfahren für das Screening eines Subjekts *in vitro*, um nachzuweisen, ob dieses Subjekt ein CF-Träger oder ein CF-Patient ist, die folgenden Schritte umfassend:
Bereitstellen einer biologischen Probe des Subjekts, das gescreent werden soll;
und Bereitstellen eines *in vitro*-Assays, ausgewählt aus einem genetischen Assay oder einem Immunoassay, für den Nachweis der Anwesenheit in der biologischen Probe von mindestens einem Glied der Gruppe bestehend aus einem mutierten CF-Gen, einem mutierten CFTR-Polypeptidprodukt, wie in Fig. 1 gezeigt, oder ihren Mischungen, wobei die Mutanten durch Mutationen an den Proteinpositionen 85, 148, 178, 455, 493, 507, 542, 549, 560, 563, 574, 1077 und 1092 definiert sind, wobei es sich bei der Änderung an Position 549 entweder um S549R oder S549I handelt.



10. Verfahren für das *in vitro*-Screening eines Subjekts, um zu bestimmen, ob das Subjekt ein CF-Träger oder ein CF-Patient ist, die folgenden Stufen umfassend:
Bereitstellen einer biologischen Probe des Subjekts, das gescreent werden soll;
und Bereitstellen eines *in vitro*-Assays, ausgewählt aus genetischem Assay und
5 Immunoassay, zum Nachweisen der Anwesenheit mindestens entweder eines mutierten CF-Gens, eines mutierten CFTR-Polypeptidprodukts oder einer ihrer Mischungen in der biologischen Probe, wobei die Mutanten durch Mutationen an den DNA-Sequenzpositionen 129, 556 und 3659 der Fig. 1, oder Mutationen in der Sequenz der Introns 4, 5 und 10, wie in Fig. 18 gezeigt, an den Nucleotidpositionen der Fig. 1, die mit 621+1, 711+1 und 1717-1 bezeichnet sind, definiert sind.
11. Nicht-menschliches Lebewesen, umfassend ein heterologes Zellsystem, umfassend einen rekombinanten Klonierungsvektor nach Anspruch 5, der cystische Fibrose-Symptome in diesem Lebewesen induziert.
12. Verwendung von Oligonucleotidprimern in einer Polymerase-Kettenreaktion (PCR), um ein selektiertes Exon zu amplifizieren, das eine cDNA-Sequenz nach Fig. 1 hat, wobei die Primer sich an den Intronabschnitten in der Nähe der 5'- und
20 3'-Grenzen eines selektierten Exons der Fig. 18 befinden und ausgewählt sind aus der Gruppe bestehend aus den Oligonucleotidprimern xi-5 und xi-3 der Tabelle 5, wo x für die Exonzahlen 1, 3, 4, 5, 6a, 6b, 7 - 13, 14a, 14b, 15, 16, 17a, 17b und 18 - 24 steht.

25



1 AATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCA
61 GAGTAGTAGGTCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGC
M Q R S P L E K A S V V S K L F 16
121 GCCCGAGAGACCATGCAGAGGTCCCTCTGGAAAAGGCCAGCGTTGTCTCCAAACTTTTT
F S W T R P I L R K G Y R Q R L E L S D 36
181 TTCACCTGGACCAGACCAATTTGAGGAAAGGATACAGACAGCGCCTGGAATTGTCAGAC
I Y Q I P S V D S A D N L S E K L E H E 56
241 ATATACCAAATCCCTTCTGTTGATTCTGCTGACAATCTATCTGAAAAATGGAAAAGAGAA
W D R E L A S K K N P K L I N A L R R C 76
301 TGGGATAGAGAGCTGGCTTCAAAGAAAAATCCTAAACTCATTAAATGCCCTTCGGCGATGT
F F W R F M F Y G I F L Y L G E V T K A 96
361 TTTTTCTGGAGATTTATGTTCTATGGAATCTTTTTATATTTAGGGGAAGTCACCAAAGCA
V O P L L L G R I I A S Y D P D N K E E 116
421 GTACAGCCTCTCTACTGGGAAGAATCATAGCTTCTATGACCCGGATAACAAGGAGGAA
R S I A I Y L G I G L C L L F I V R T L 136
481 CGCTCTATCGCGATTTATCTAGGCATAGGCTTATGCCTTCTCTTTATTTGTGAGGACACTG
L L H P A I F G L H H I G H Q M R I A M 156
541 CTCCTACACCCAGCCATTTTTGGCCTTCATCACATTGGAATGCAGATGAGAATAGCTATG
F S L I Y K K T L K L S S R V L D K I S 176
601 TTTAGTTTGATTTATAAGAAGACTTTAAAGCTGTCAAGCCGTGTTCTAGATAAAAATAAGT
I G Q L V S L L S N N L N K F D E G L A 196
661 ATTGGACAACCTGTTAGTCTCCTTTCCAACAACCTGAACAAATTTGATGAACGACTTGCA
L A H F V W I A P L O V A L L M G L I W 216
721 TTGGCACATTTCTGTTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGGCTAATCTGG
E L L Q A S A F C G L G F L I V L A L E 236
781 GAGTTGTTACAGGGCTCTGCCTTCTGTGGACTTGGTTTCTGATAGTCCTTGCCTTTTT
O A G L G R M M M K Y R D Q R A G K I S 256
841 CAGGCTGGGCTAGGGAGAATGATGATGAAGTACAGAGATCAGAGAGCTGGGAAGATCAGT
E R L V I T S E M I E N I Q S V K A Y C 276
901 GAAAGACTTGTGATTACCTCAGAAATGATTGAAAATATCCAATCTGTTAAGGCATACTGC
W E E A M E K M I E N L R C T E L K L T 296
961 TGGGAAGAAGCAATGGAAAAATGATTGAAAATTAAGACAACAGAACTGAAACTGACT
R K A A Y V R Y F N S S A F F F S G F F 316
1021 CGGAAGGCAGCCTATGTGAGATACTCAATAGCTCAGCCTTCTTCTTCTCAGGGTTCTTT
V V F L S V L P Y A L I K G I I L R K I 336
1081 GTGGTGTTTTTATCTGTGCTTCCCTATGACTAATCAAAGGAATCATCCTCCGAAAATA
F T T I S F C I V L R M A V T R Q F P W 356
1141 TTCACCACCATCTCATTCTGCATTGTTCTGCGCATGGCGGTCACTCGGCAATTTCCCTGG
A V Q T W Y D S L G A I N K I Q D F L Q 376
1201 GCTGTACAAACATGGTATGACTCTCTTGGAGCAATAAACAAAATACAGGATTTCTTACAA
K Q E Y K T L E Y N L T T T E V V M E N 396
1261 AAGCAAGAATATAAGACATTGGAATATAACTTAACGACTACAGAAGTAGTGATGGAGAAT
V T A F W E E G F G G E L F E K A K Q N N 416
1321 GTAACAGCCTTCTGGGAGGAGGATTTGGGGAATTTATTTGAGAAAGCAAACAACAACAAT

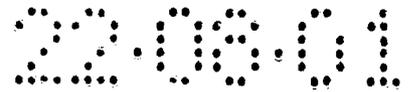


FIG. 1 (Fortsetzung)

2145

1381 N N R K T S N G D D S L F F S N F S L L 436
AACAAATAGAAAACTTCTAATGGTGATGACAGCCTCTTCTTCAGTAATTTCTCACTTCTT

1441 G T P V L K D I H F K I E R G Q L L A V 456
GGTACTCCTGTCCTGAAAGATATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGCGGTT

1501 A G S T G A G K T S L L M M I M G E L E 476
GCTGGATCCACTGGAGCAGGCAAGACTTCACTTCTAATGATGATTATGGGAGAACTGGAG

1561 P S E G K I K E S G R I S F C S Q F S W 496
CCTTCAGAGGGTAAAATTAAGCACAGTGGGAAGAATTTCACTTCTGTTCTCAGTTTTCTGG

1621 I M P G T I K E H I I F G V S Y D E Y R 516
ATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTCCTATGATGAATATAGA

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 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
TCTTTAGCAAGAGCAGTATACAAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGG

1861 Y L D V L T E K E I F E S C V C K L M A 596
TACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAGCTGTGTCTGTAAACTGATGGCT

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 H E G S S Y F Y G T F S E L Q N L 636
TTAATTTTGCATGAAGGTAGCAGCTATTTTTATGGGACATTTTCAGAACTCCAAAATCTA

2041 Q P D F S S K L M G C D S F D Q F S A E 656
CAGCCAGACTTTAGCTCAAAAACCTCATGGGATGTGATTCTTTGACCAATTTAGTGAGAA

2101 R R N S I L T E T L H R F S L E G D A P 676
AGAAGAAATTCATCTCAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCT

2161 V S W T E T K K Q S F K Q T G E F G E K 696
GTCTCCTGGACAGAAACAAAAACAATCTTTTAAACAGACTGGAGAGTTGGGGAAAAA

2221 R K N S I L N P I N S I R K F S I V Q K 716
AGGAAGAATCTATTCTCAATCCAATCAACTCTATACGAAAATTTCCATTGTGCAAAAG

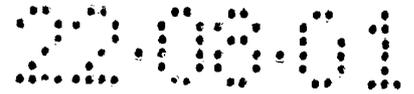
2281 T P L Q M N G I E E D S D E P L E R R L 736
ACTCCCTTACAAATGAATGGCATCGAAGAGGATTCTGATGAGCCTTTAGAGAGAAGGCTG

2341 S L V P D S E Q 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 Q S V L N L M T H S 776
ACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGTCTGAACTGATGACACACTCA

2461 V N Q G Q N I H R K T T A S T R K V S L 796
GTTAACCAAGGTGAGAACATTCACCGAAAGACAACAGCATCCACACGAAAAGTGTCACTG

2521 A P Q A N L T E L D I Y S R R L S Q E T 816
GCCCCTCAGGCAAACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAACT



2581 G L E I S E E I N E E D L K | E C F F D D 836
 GGCTTGAAATAAGTGAAGAAATTAACGAAGAAGACTTAAAGGAGTGC^TTTTTTGATGAT

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

2701 K S L I F V L I W C L V I F L A E V A A 876
 AAGAGCTTAATTTTTGTGCTAATTTGGTGCTTAGTAATTTTCTGGCAGAG^TTTGGCTGCT

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

2821 H S R N N S Y A V I I T S T S S Y Y 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
 TACATTTACGTGGGAGTAGCCGACACTTTGCTTGCTATGGGATTCTTCAGAGGTCTACCA

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

3001 L Q A P M S T L N T L K A ^{17A} E G I L N R F 976
 CTTCAAGCACCTATGTCAACCCTCAACACGTTGAAAGCAGTGGGATTCTTAATAGATTC

3061 S K D I A I L D D L L P L T I F D F I Q 996
 TCCAAAGATATAGCAATTTTGGATGACCTTCTGCCTCTTACCATATTTGACTTCATCCAG

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

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

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

3301 T S L K G L W T L R A F G R Q P Y F E T 1076
 ACAAGCTTAAAGGACTATGGACACTTCGTGCCTTCGGACGGCAGCCTTACTTTGAAACT

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

3421 R W F Q M R I E M I F V I F F I A V T F 1116
 CGCTGGTTCCAAATGAGAATAGAAATGATTTTGTGCTCTTCTTCATTGCTGTTACCTTC

3481 I S I L T T G E G E G F V G I I L T L A 1136
 ATTTCCATTTTAAACAACAGGAGAAGGAGAAGGAAGAGTTGGTATTATCCTGACTTTAGCC

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

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

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

3721 H V K K D D I W P S G G Q 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
 GCAAAAATACACAGAAGGTGGAAATGCCATATTAGAGAACATTTCTTCTCAATAAGTCTCT

3841 G Q R V G L L G R T G S G K S T L L S A 1256
 GGCCAGAGG^TGTGGGCTCTTGGGAAGAACTGGATCAGGGAAGAGTACTTTGTTATCAGCT

3901 F L R L L N T E G E I Q I D G V S W D S 1276
 TTTTGGAGACTACTGAACACTGAAGGAGAAATCCAGATCGATGGTGTGTCTTGGGATTCA
 3961 I T L Q Q W R K A F G V I P Q ^{20 21} K V F I F 1296
 ATAACCTTTGCAACAGTGGAGGAAAGCCTTTGGAGTGATACCACAGAAAGTATTATTTTT
 4021 S G T F R K N L D P Y E Q W S D Q E I W 1316
 TCTGGAACATTTAGAAAAAAGCTTGGATCCCTATGAACAGTGGAGTGATCAAGAAATATGG
 4081 K V A D E V G L R S V I E Q F P G K L D 1336
 AAAGTGCAGATGAGCTTGGGCTCAGATCTGTGATAGAACAGTTTCTGGGAAGCTTGAC
 4141 F V L V D G G C V L S E G H K Q L M C L 1356
 TTTGTCCTTGTGGATGGGGCTGTGTCCTAAGCCATGGCCACAAGCAGTTGATGTGCTTG
 4201 A R S V L S K A K I L L L D E P S A H L 1376
 GCTAGATCTGTTCTCAGTAAGGCGAAGATCTTGCTGCTTGATGAACCCAGTGCTCATTGG
 4261 D P V T Y Q I I R R T L K Q A F A D C T 1396
 GATCCAGTAACATACCAAATAATTAGAAGAACTCTAAAACAAGCATTGCTGATTGCACA
 4321 V I L C E H R I E A M L E C Q Q F L V I 1416
 GTAATTCTCTGTGAACACAGGATAGAAGCAATGCTGGAATGCCAACAATTTTTGGTCATA
 4381 E E N K V R Q Y D S I Q K L L N E R S L 1436
 GAAGAGAACAAGTGGGGCAGTACGATTCCATCCAGAAACTGCTGAACGAGAGGAGCCTC
 4441 F R Q A I S P S D R V K L F P H R N S S 1456
 TTCCGGCAAGCCATCAGCCCCTCCGACAGGGTGAAGCTCTTCCCCACCGGAAGCTCAAGC
 4501 K C K S K P Q I A A L K E E T E E E V Q 1476
 AAGTGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGAGACAGAAGAAGAGGTGCAA
 D T R L = 1480
 4561 GATACAAGGCTTTAGAGAGCAGCATAAATGTTGACATGGGACATTTGCTCATGGAATTGG
 4621 AGCTCGTGGGACAGTCACCTCATGGAATTGGAGCTCGTGGACAGTTACCTCTGCCTCAG
 4681 AAAACAAGGATGAATTAAGTTTTTTTTTAAAAAAGAAACATTTGGTAAGGGGAATTGAGG
 4741 ACCTGATATGGGTCTTGATAAATGGCTTCCCTGGCAATAGTCAAATTGTGTGAAAGGTAC
 4801 TTCAAATCCTTGAAGATTTACCACTTGTGTTTTGCAAGCCAGATTTTCTGAAAACCCTT
 4861 GCCATGTGCTAGTAATTGGAAGGCAGCTCTAAATGTCAATCAGCCTAGTTGATCAGCTT
 4921 ATTGCTAGTGAAACTCGTTAATTTGTAGTGTGGAGAAGAAGTAAATCATACTTCTTA
 4981 GGGTTATGATTAAGTAATGATAACTGGAAGCTTCCAGCGGTTTATATAAGCTTGTATTCCCT
 5041 TTTTCTCTCCTCTCCCCATGATGTTTAGAAACACAAGTATATTGTTTGGTAAAGCATTCCA
 5101 ACTATCTCATTTCGAAGCAAGTATTAGAATACCACAGGAACCACAAGACTGCACATCAAA
 5161 ATATGCCCCATTCAACATCTAGTGAGCAGTCAGGAAAGAGAAGTCCAGATCCTGGAAAT
 5221 CAGGGTTAGTATTGTCCAGGCTACCAAAAATCTCAATATTTAGATAATCACAATACAT
 5281 CCCCTACCTGGGAAAGGGCTGTTATAATCTTTCACAGGGGACAGGATGGTCCCTTGATG
 5341 AAGAAGTTGATATGCCTTTTCCCAACTCCAGAAAAGTGACAAGCTCACAGACCTTTGAACT
 5401 AGAGTTTAGCTGGAAAAGTATGTTAGTGCAAATTGTACAGGACAGCCCTTCTTTCCACA
 5461 GAAGCTCCAGGTAGAGGGTGTGTAAGTAGATAGGCCATGGGCACTGTGGGTAGACACACA
 5521 TGAAGTCCAAGCATTAGATGTATAGGTTGATGGTGGTATGTTTTAGGCTAGATGTATG
 5581 TACTTCATGCTGTCTACACTAAGAGAGAATGAGAGACACACTGAAGAAGCACCATCATG
 5641 AATTAGTTTTATATGCTTCTGTTTTATAATTTGTGAAGCAAAATTTTTCTCTAGGAAA
 5701 TATTTATTTTAAATAATGTTTCAAACATATATTACAATGCTGTATTTAAAAGAAATGATTA
 5761 TGAATTACATTTGTATAAAATAATTTTTATATTTGAAATATTGACTTTTTATGGCACTAG
 5821 TATTTTTATGAAATATTATGTTAAAAGTGGGACAGGGGAGAACCTAGGGTGATATTAACC
 5881 AGGGGCCATGAATCACCTTTTGGTCTGGAGGGGAGCCCTTGGGGCTGATCGAGTTGTGGCC
 5941 CACAGCTGTATGATTTCCAGCCAGCAGCCCTCTTAGATGCAGTTCTGAAGAAGATGGT
 6001 ACCACCAGTCTGACTGTTTCCATCAAGGGTACACTGCCTTCTCAACTCCAAGTACTCT
 6061 TAAGAAGACTGCATTATATTTACTGTAAGAAAATATCACTTGTCAATAAATCCATA
 6121 CATTGTGT (A) n

FIG. 2 (Fortsetzung)

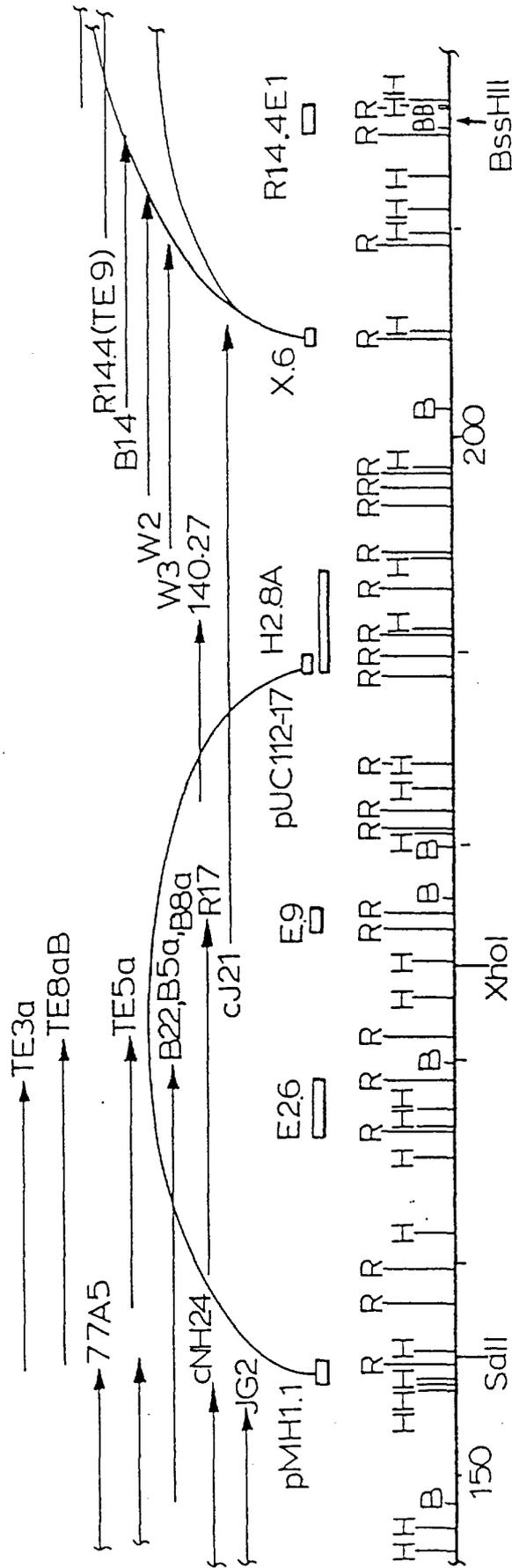


FIG. 2 (Fortsetzung)

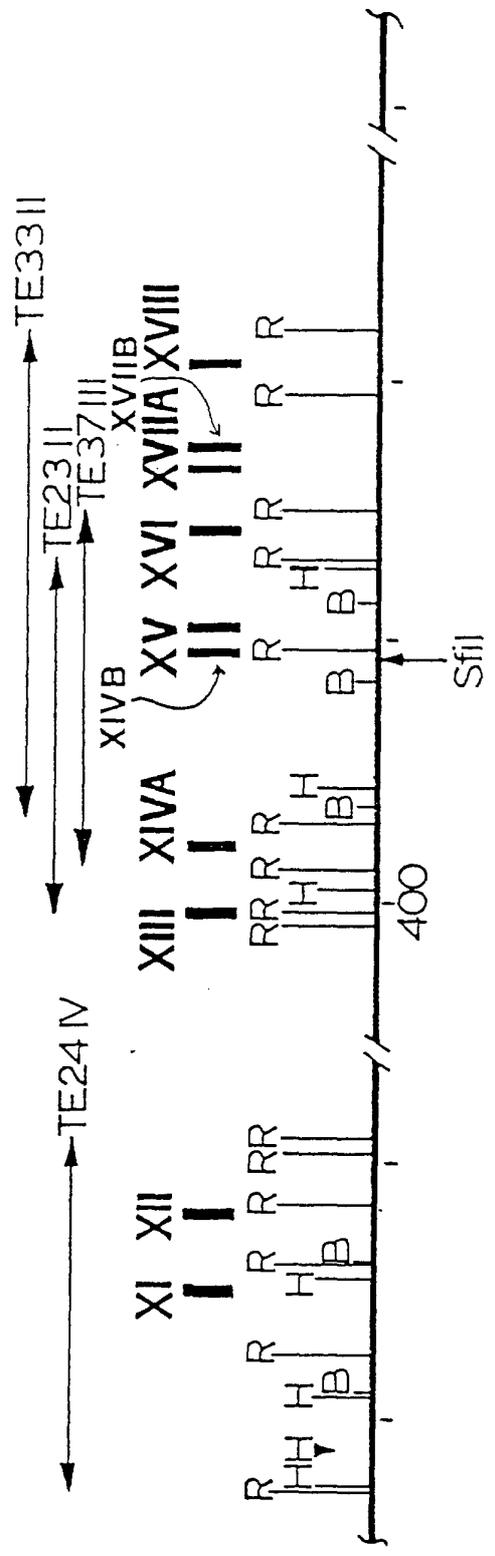


FIG. 2 (Fortsetzung)



11/45

12/45

FIG. 3A FIG. 3B FIG. 3C FIG. 3D

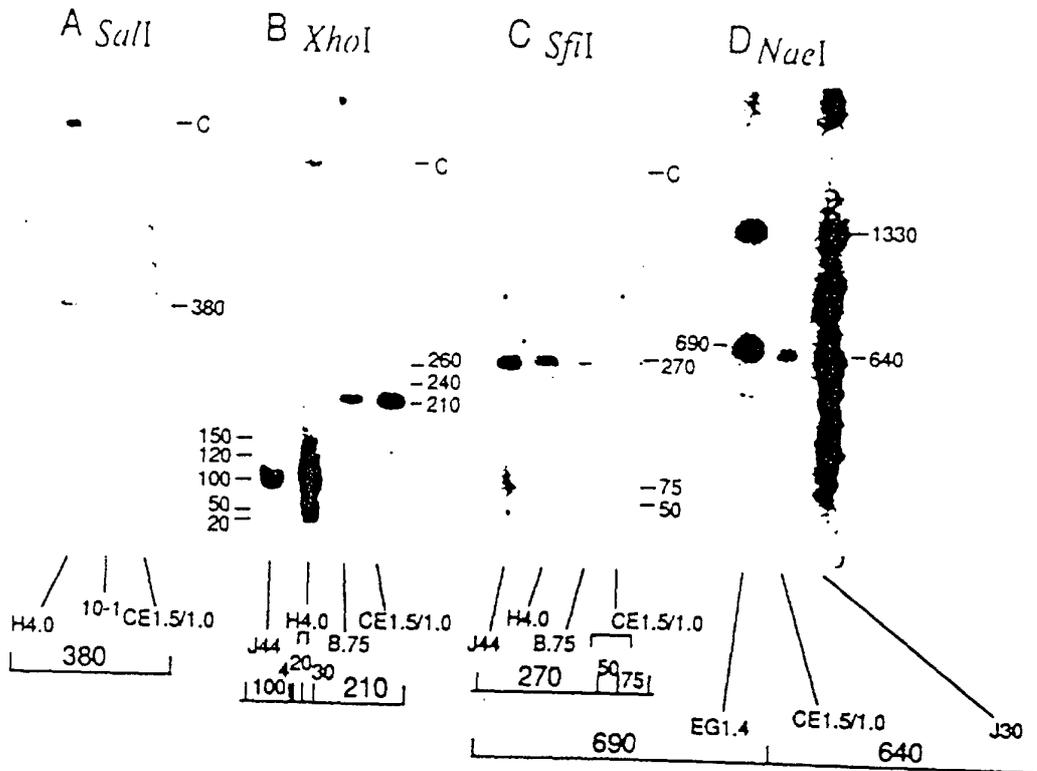
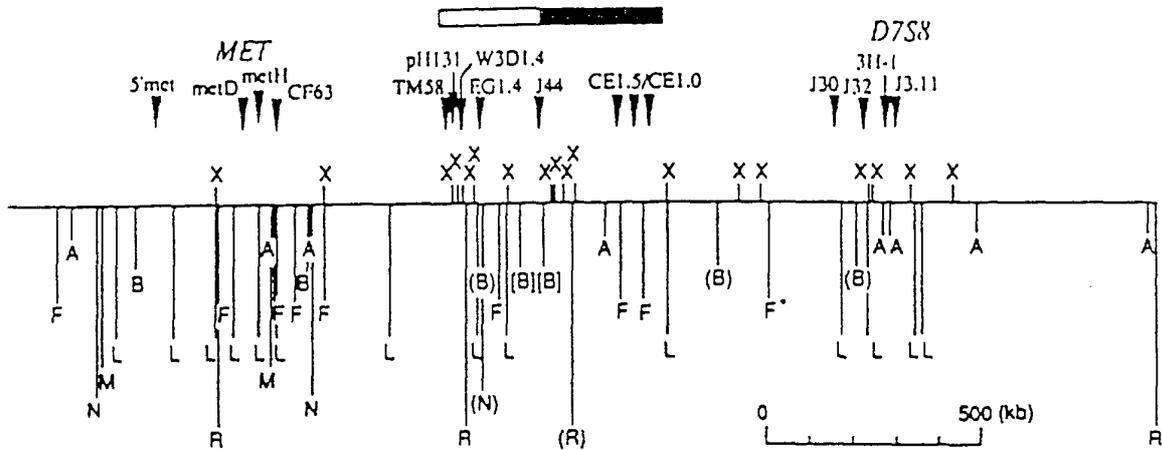


FIG. 3E



14/45

FIG. 4A

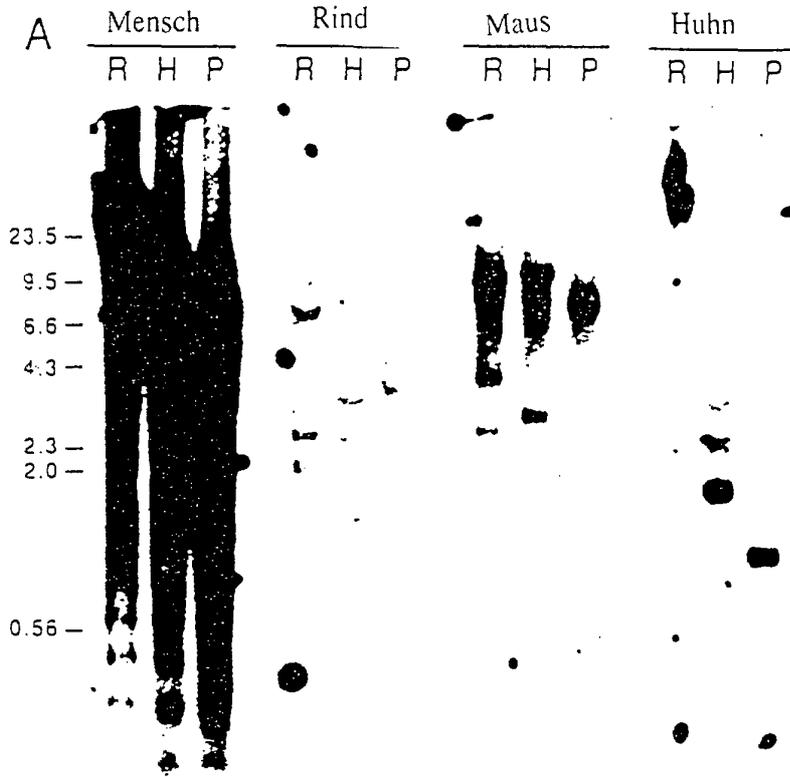
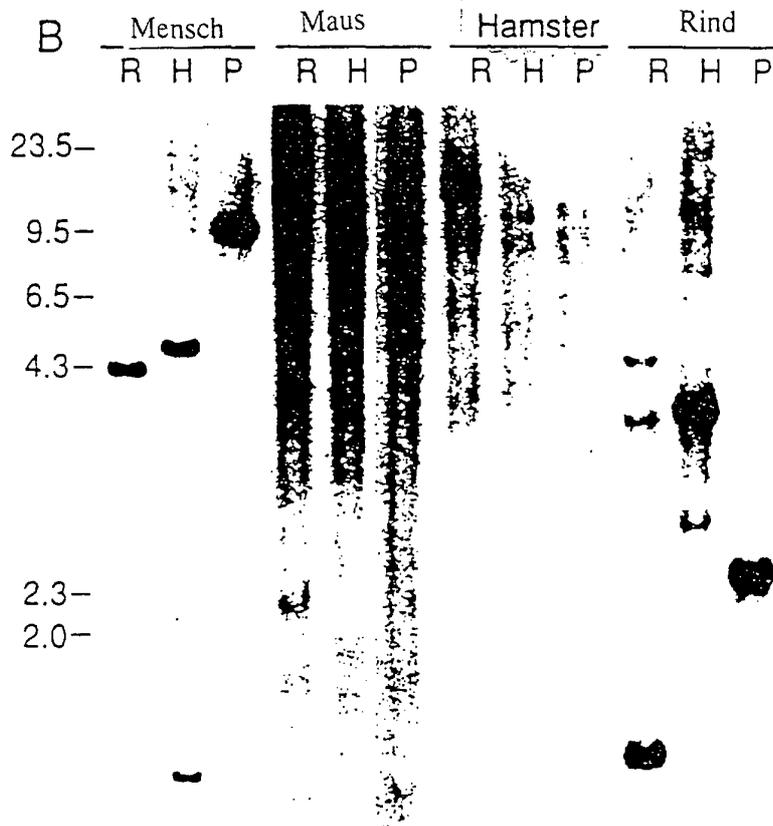


FIG. 4B



15/45

FIG. 4C

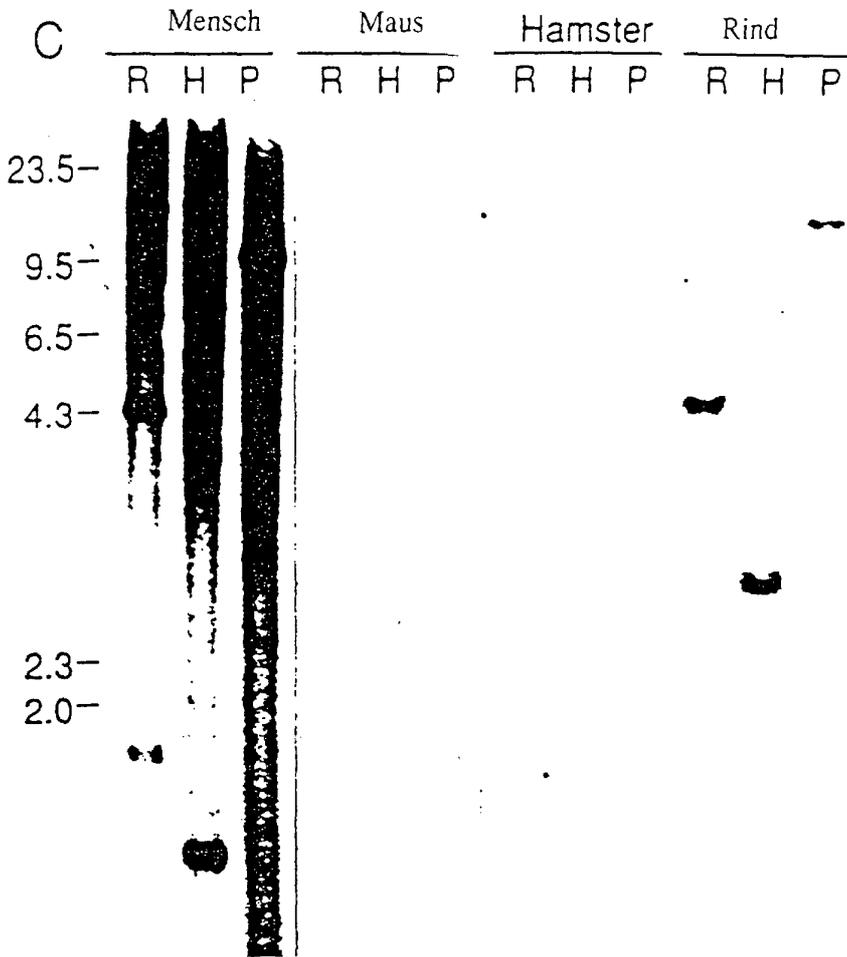
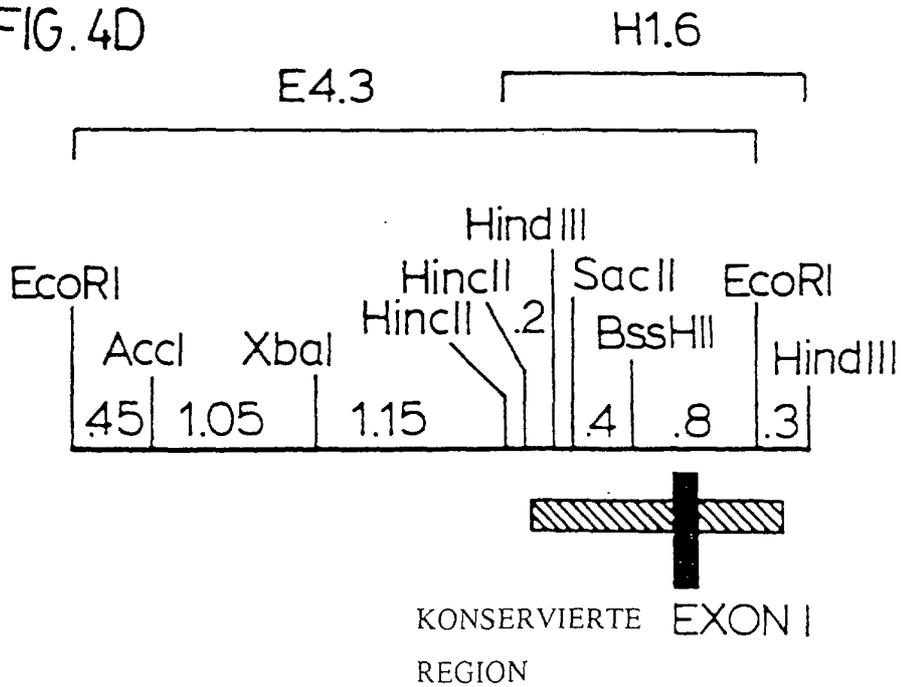


FIG. 4D



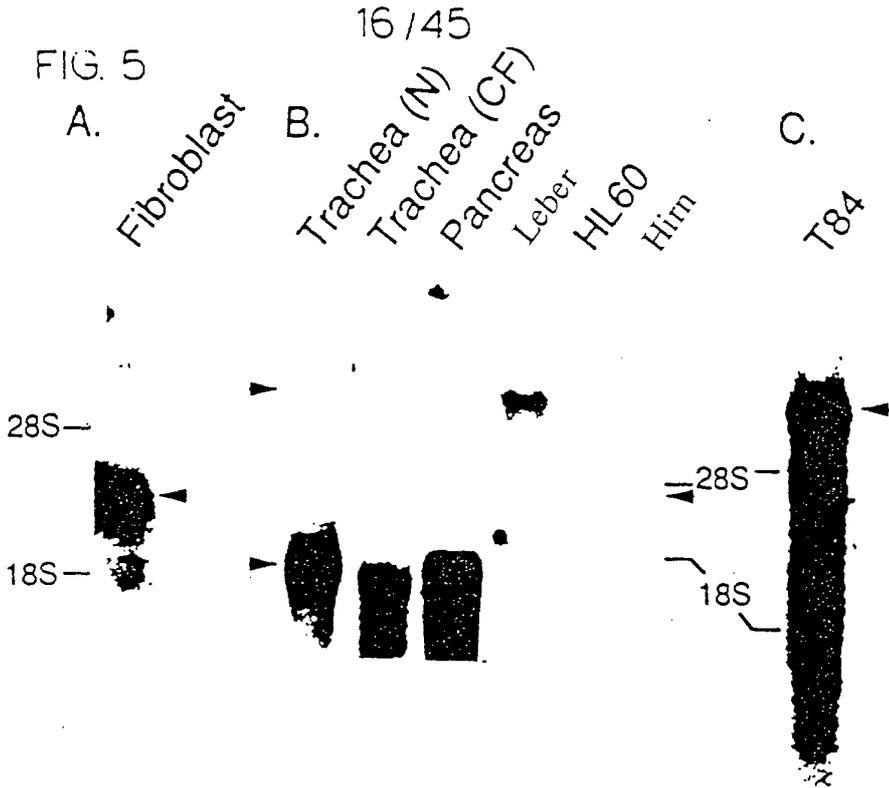


FIG. 6



FIG. 7.

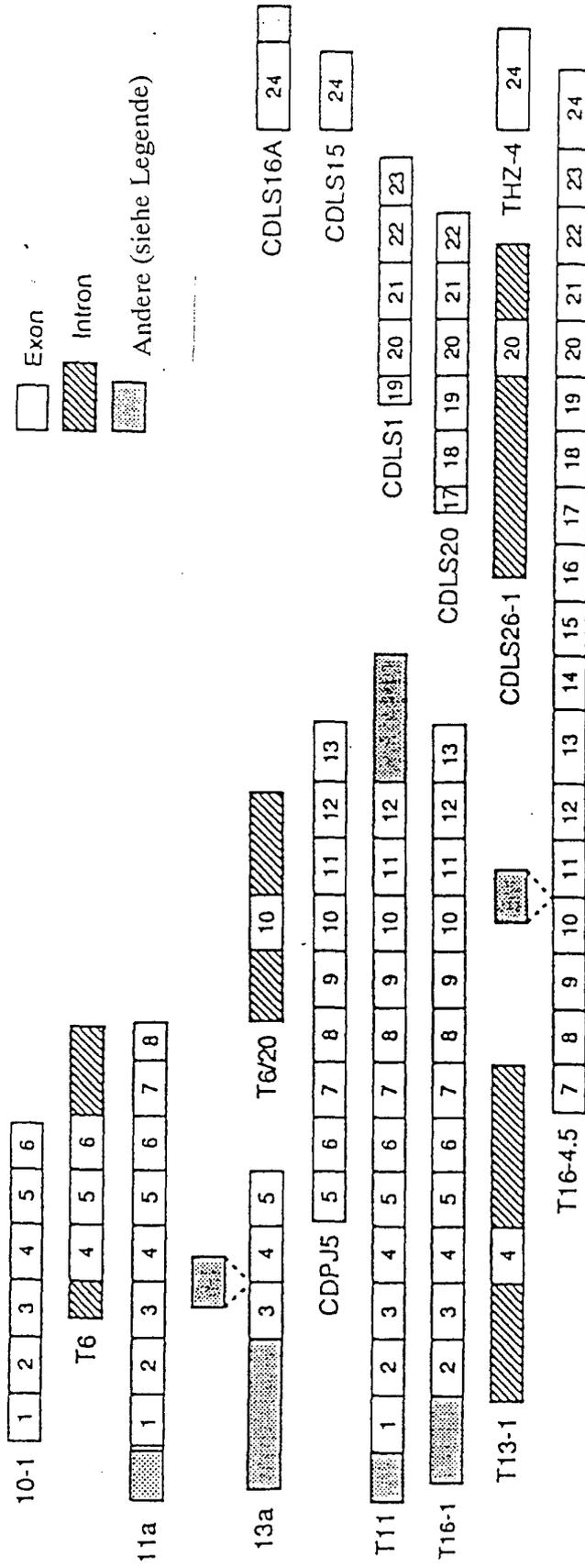
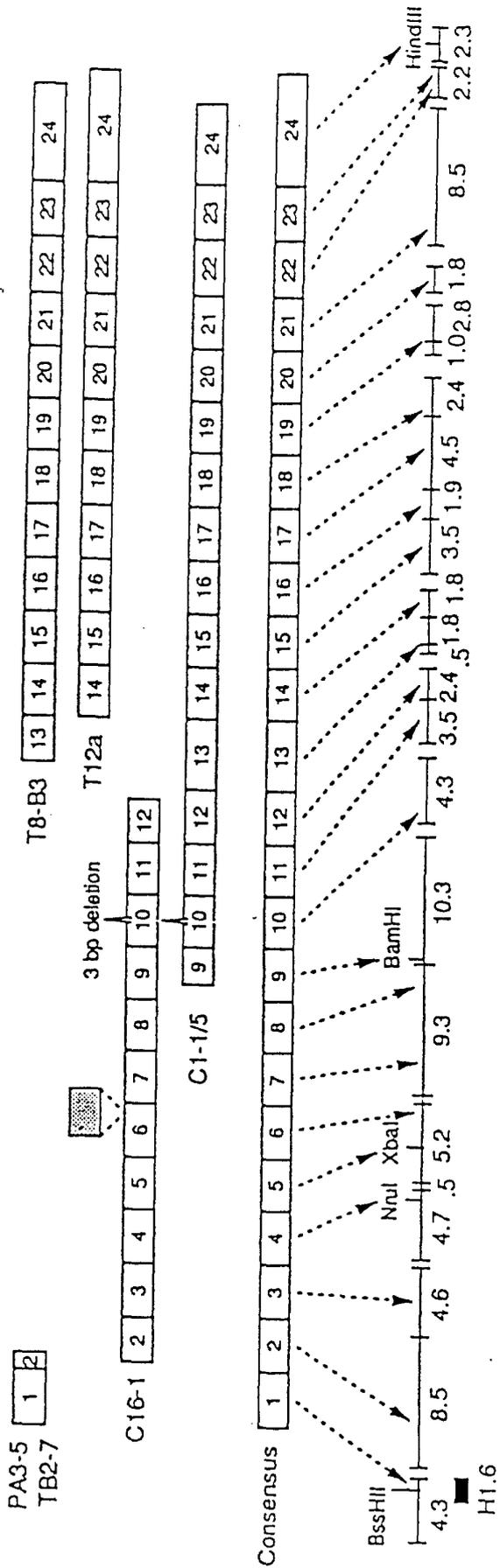
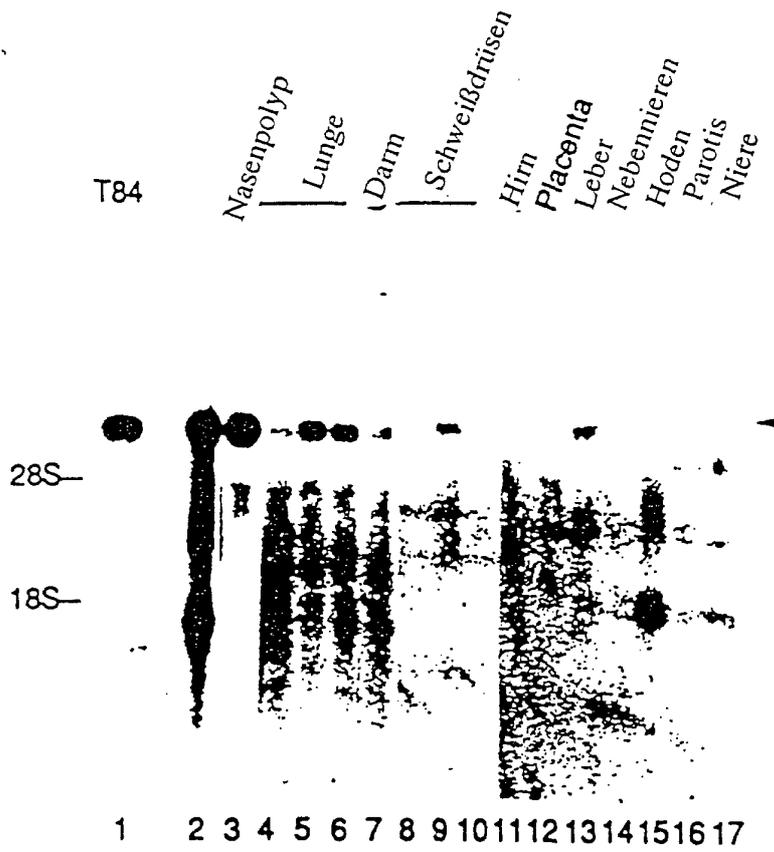


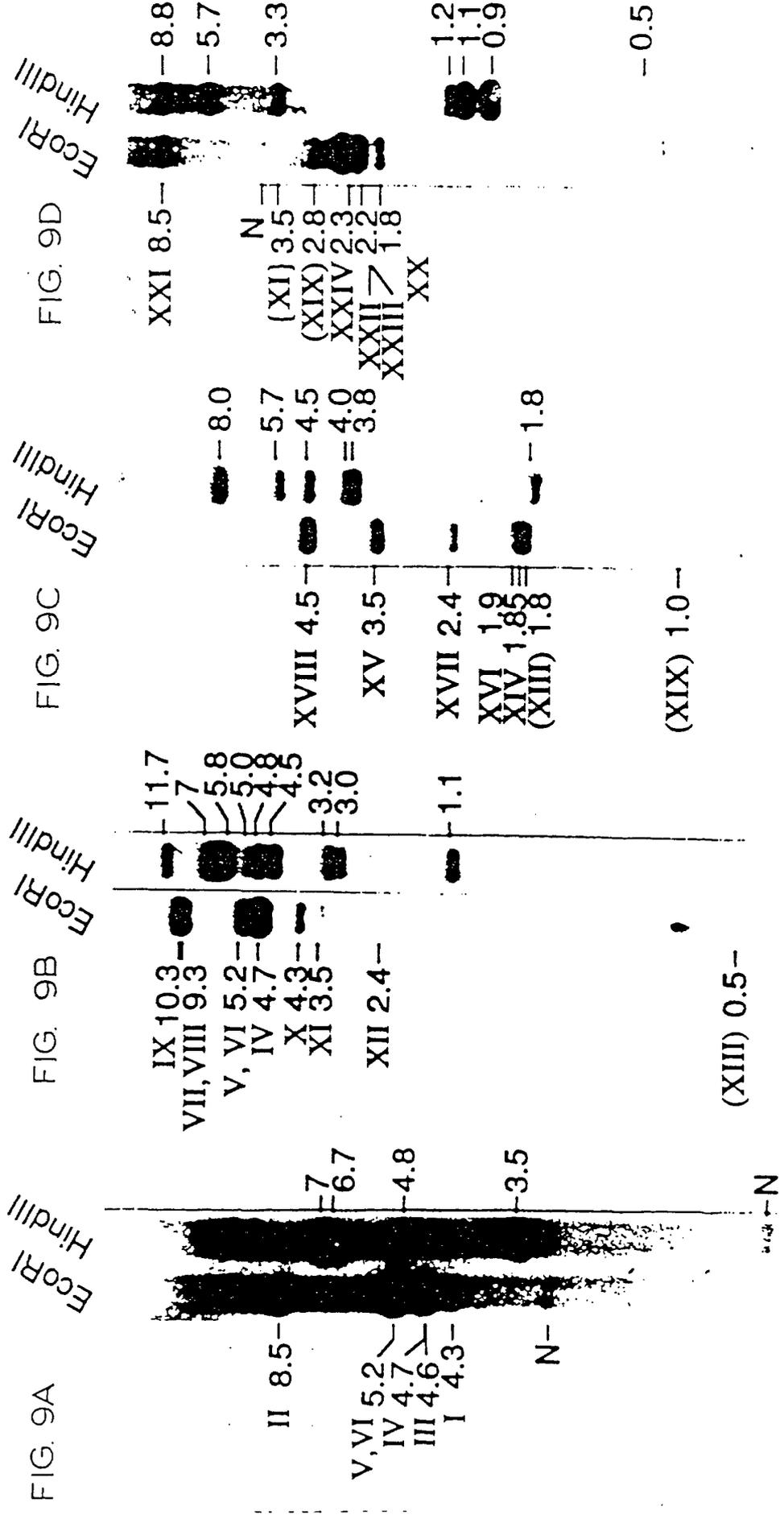
Fig. 7 (Fortsetzung)



19/45

FIG. 8

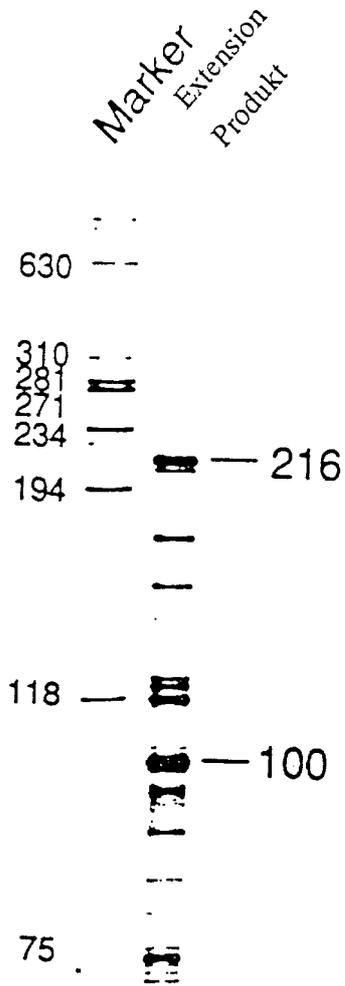




—1.5

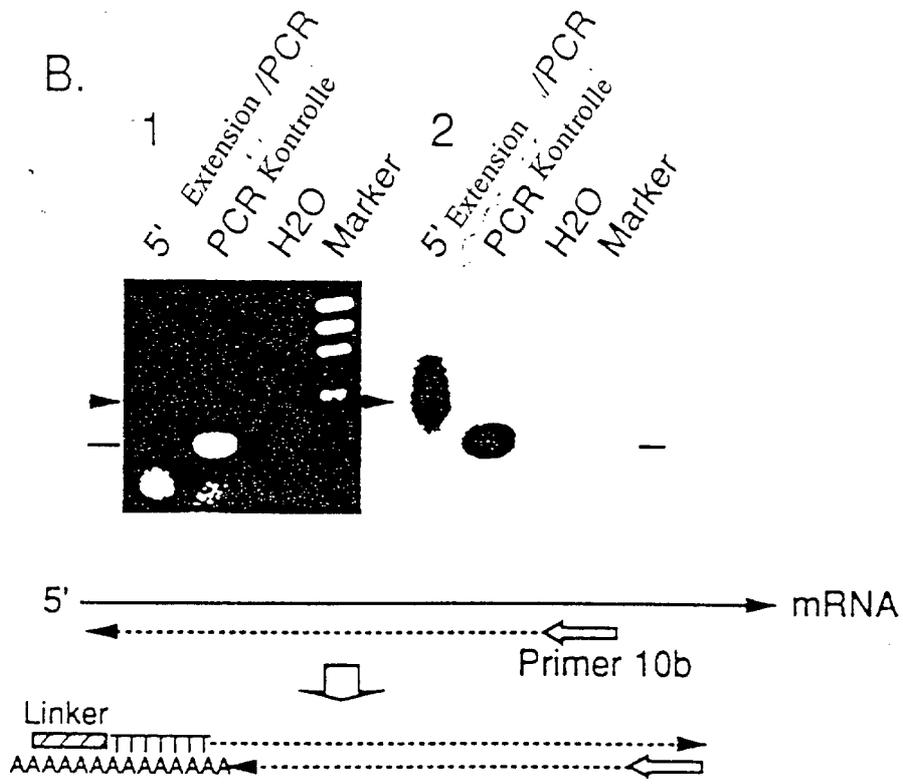
21/45

FIG. 10A



22/45

FIG. 10B



23/45

FIG. 10C

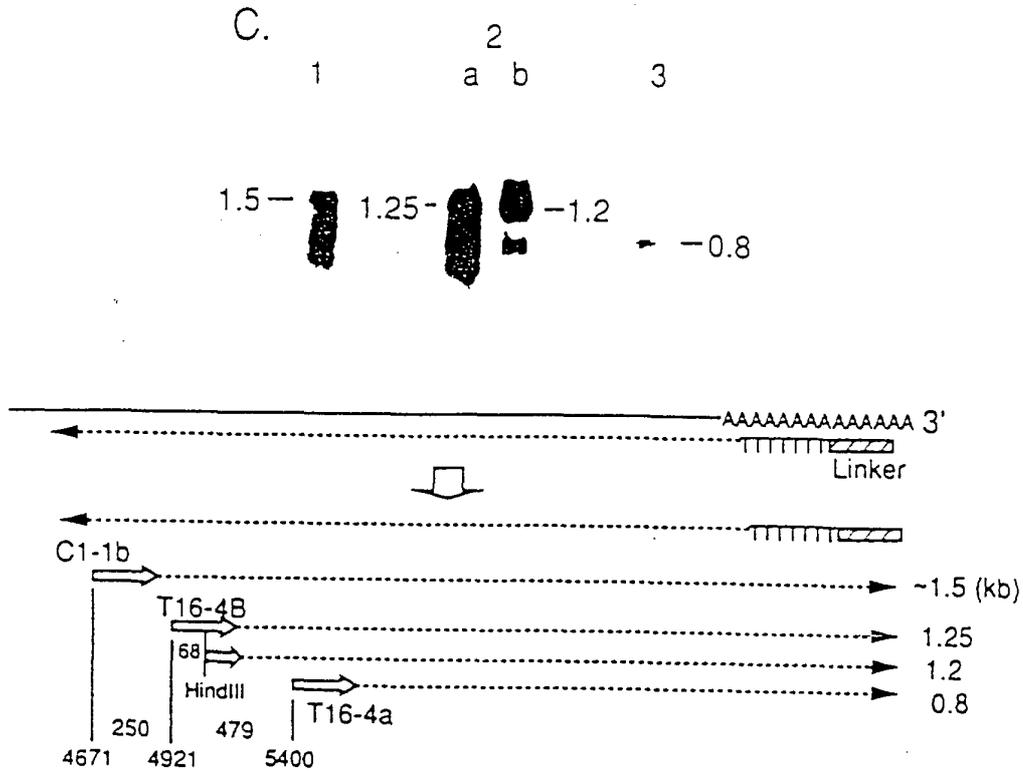


FIG. 11.

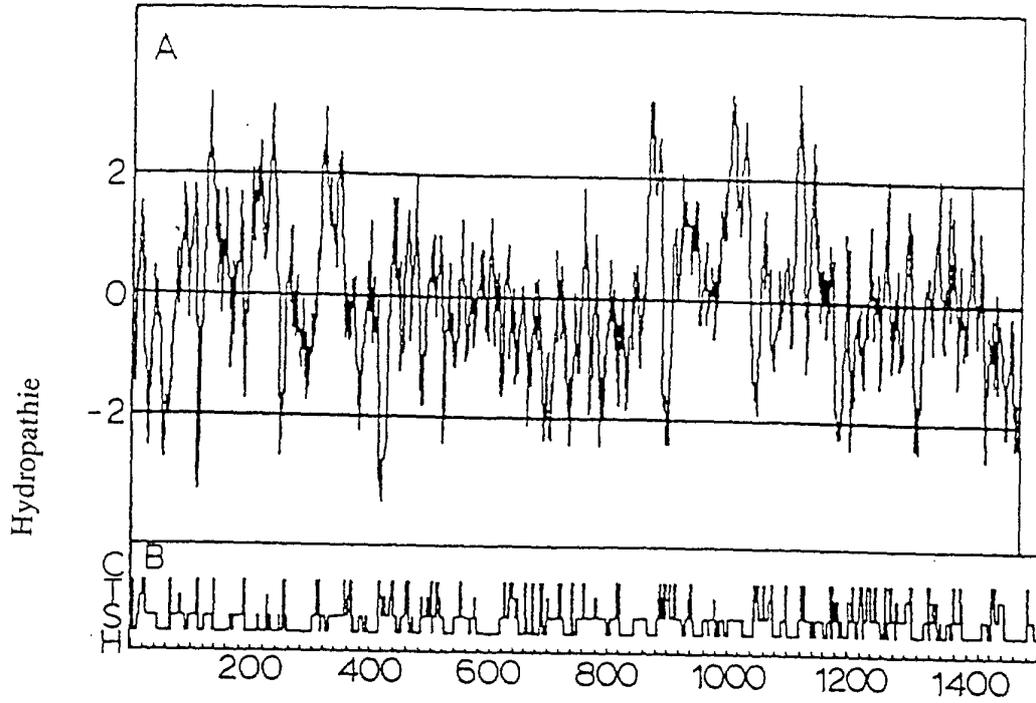


FIG. 12

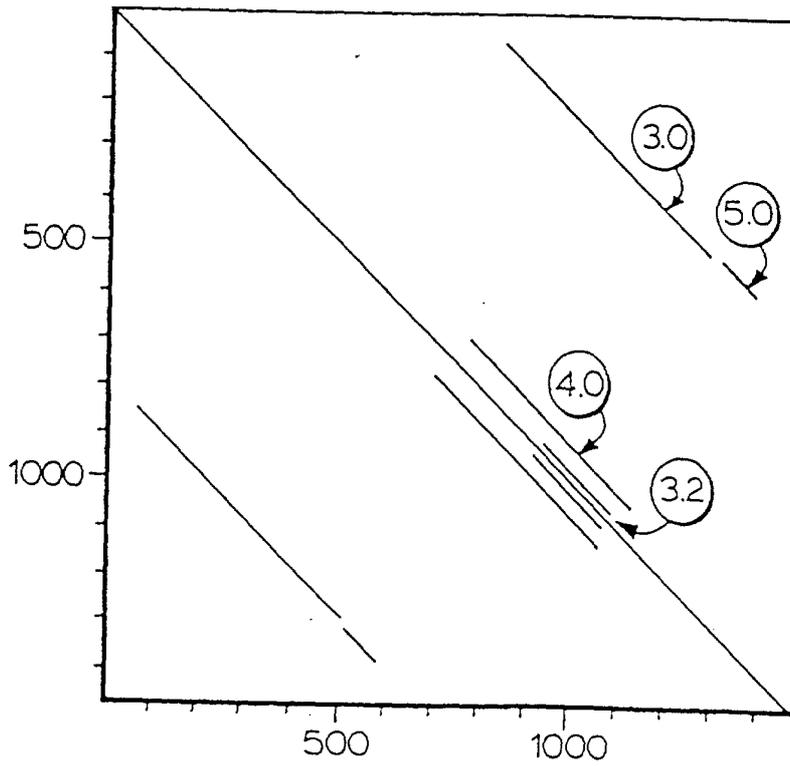
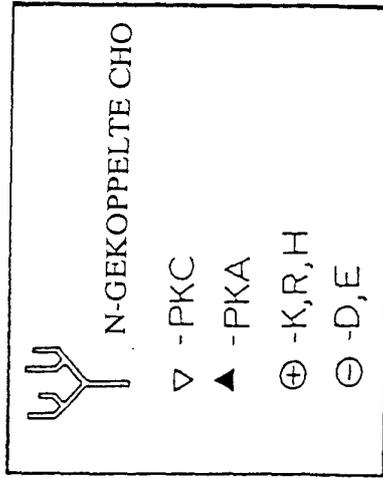
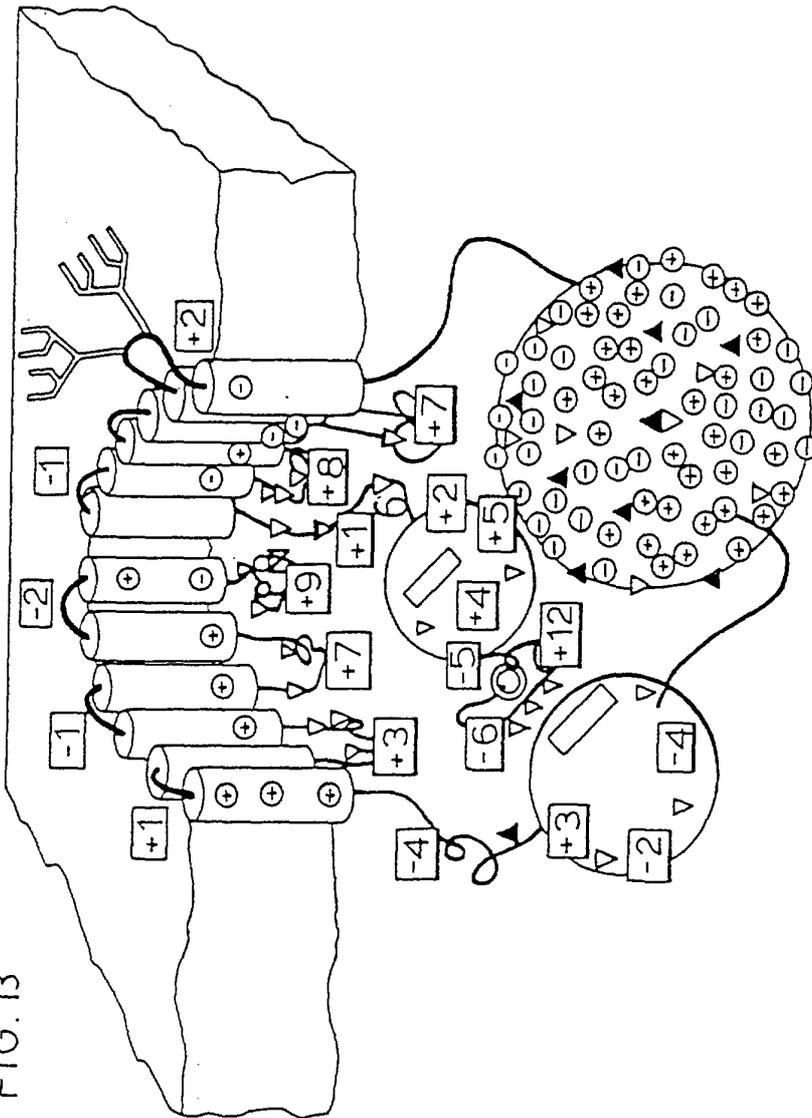


FIG. 13



CFTR (N)	F S L L G T P V L K D I N F K I E R G Q L L A V A G S T G A G K T S L L M M I M G
CFTR (C)	Y T E G N A I L E N I S F S I S P G Q R V G L L G R T G S G K S T L L S A F L R
hmdr1 (N)	P S R K E V K I L K G L N L K V Q S G Q T V A L V G N S G C G K S T T V Q L M Q R
hmdr1 (C)	P T R P D I P V L Q G L S L E V K K G Q T L A L V G S S G C G K S T V V Q L L E R
mmdr1 (N)	P S R S E V Q I L K G L N L K V K S G Q T V A L V G N S G C G K S T V V Q L M Q R
mmdr1 (C)	P T R P N I P V L Q G L S L E V K K G Q T L A L V G S S G C G K S T V V Q L L E R
mmdr2 (N)	P S R A N I K I L K G L N L K V K S G Q T V A L V G N S G C G K S T T V Q L L Q R
mmdr2 (C)	P T R A N V P V L Q G L S L E V K K G Q T L A L V G S S G C G K S T V V Q L L E R
p fmdr (N)	D T R K D V E I Y K D L S F T L L K E G K T Y A F V G E S G C G K S T I L K L I E
p fmdr (C)	I S R P N V P I Y K N L S F T C D S K K T T A I V G E T G S G K S T F M N L L L R
STE6 (N)	P S R P S E A V L K N V S L N F S A G Q T F I V G K S G S G K S T L S N L L L R
STE6 (C)	P S A P T A F V Y K N M N F D M F C G Q T L G I I G E S G T G K S T L V L L L T K
hlyB	Y K P D S P V I L D N I N I S I K Q G E V I G I V G R S G S G K S T L I K L I Q R
White	I P A P R K H L L K N V C G V A Y P G E L L A V M G S S G A G K T L L L N A L A F
MbpX	K S L G N L K I L D R V S L Y V P K F S L I A L L G P S G S G K S L L R I L A G
Btud	Q D V A E S T R L G P L S G E V R A G R I L H L V G P N G A G K S T L L A R I A G
PstB	F Y Y G K F H A L K N I N L D T A K N Q V T A F I G P S G C G K S T L L R T F N K
hisp	R R Y G G H E V L K G V S L Q A R A G D V I S I I G S S G S G K S T F L R C I N F
malK	K A W G E V V V S K D I N I D I H E G E F V V F V G P S G C G K S T L L R M I A G
oppD	T P D G D V T A V N D L N F T L R A G E T L G I V G E S G S G K S Q T A F A L M G
oppF	Q P P K T L K A V D G V T L R L Y E G E T L G V V G E S G C G K S T F A R A I I G
RbsA (N)	K A V P G V K A L S G A L N V Y P G R V M A L V G E N G A G K S T M M K V L T G
RbsA (C)	V D N L C G P G V N D V S F T L R K G E I L G V S G L M G A G R T E L M K V L Y G
UvrA	L T G A R G N N L K D V T L T L P V G L F T C I T G V S G S G K S T L I N D T L F
NodI	K S Y G G K I V V N D L S F T I A A G E C F G L L G P N G A G K S T I I R M I L G
Ftse	A Y L G G R Q A L Q G V T F H M Q P G E M A F L T G H S G A G K S T L L K L I C G

*

ISFCSQFSWIMPGTIK-ENI IFGVSYD
DSITLQQRKAFGVIPQKVFIFSGTFR
IGVVSQEPVLFATTI-AENIRYGRNV
LGIVSQEPILFDCSI-AENIAYGDNRSR
IGVVSQEPVLFATTI-AENIRYGREDDV
LGEVSQEPILFDCSI-AENIAYGDNRSR
IGVVSQEPVLSFTTI-AENIRYGRGNV
LGIVSQEPILFDCSI-AENIAYGDNRSR
IGVVSQDPLLEFSNSI-KNNIKYSLSL
FSIVSQEPMLEFMSI-YENIKFGREDA
ITVVEQRCTLFNDTL-RKNILLGSTDS
ISVVEQKPLLEFNGTI-RDNLTYGLQDE
VGVVLDQDNVLLNRSI-IDNISLAPGMS
RCAYVQQDDLFIGLIAREHLIFQAMVR
MSFVFQHYALFKHMTVYENISFGLRLR
YLSQQQTPPFATPVVWHYLLHQHDKTR
VGMVFKQPTFPFMSI-YDNIAFGVRLF
GIMVFQHFNLWSHMTVLENVMEAPIQV
VGMVFSYALYPHLSVAENMSFGLKPA
ISMIFQDPMTSLNPMYMRVGEQLMEVLM
IQMIFQDPLASLNPRMTIGEIIAEPLR
AGIIHQEELNLIPLQTLAENIFLGRFV
ISEDKRKRDGLVLGMSVKENMSLTALRY
TYTGVTTPVRELFAGVPESTRARGYTPG
IGIVSQEDNLDLEFTVRENLLVYGRYF
IGMIFQDHHLLMDRTVYDVAIPLIIA

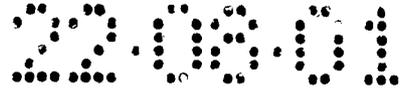


FIG. 15 (Fortsetzung)

28/45

CFTR	(N)	GEGITLSGGQRARISLARAVYKDADLYLLDSPFGYLDVLT
CFTR	(C)	VDGGCVLSHGKQMLARSVLSKAKILLDDEPSAHLDPVTYQ
hmdr1	(N)	GERGAQLSGGQKORIAIARALVRNPKILLDEATSALDTESEA
hmdr1	(C)	GDKGTLSSGGQKORIAIARALVRNPKILLDEATSALDTESEA
mmdr1	(N)	GERGAQLSGGQKORIAIARALVRNPKILLDEATSALDTESEA
mmdr1	(C)	GDKGTQLSSGGQKORIAIARALVRNPKILLDEATSALDTESEA
mmdr2	(N)	GDRGAQLSGGQKORIAIARALVRNPKILLDEATSALDTESEA
mmdr2	(C)	GDKGTQLSSGGQKORIAIARALVRNPKILLDEATSALDTESEA
pfmdr	(N)	GSNASKLSGGQKORISIAARAIMRNPKILLDEATSALDTESEA
pfmdr	(C)	PYGKS-LSSGGQKORIAIARALVRNPKILLDEATSALDTESEA
STE6	(N)	GTGGVTLSSGGQQRVAIARAFIRDTPILFLDEAVSALDIVHRN
STE6	(C)	RIDTTLSSGGQQRVAIARAFIRDTPILFLDEAVSALDIVHRN
hlyB		GEQAGLSGGQQRVAIARALVRNPKILLDEATSALDTESEA
White		PGRVKGLSSGGERKRLAFASEALDPPILLICDEPTSGLDSTAH
Mbpx		FEYPAQLSSGGQQRVALARSLAIQPDLLL-DEPFGALDDELRR
Btud		GRSTNQLSSGGEWQRVLAAVLQITLLLLDEPMNSLDVAQQA
PstB		HQSGYSLSSGGQQRVLCIARGIAIRPEVLLLDPECSALDPISTG
hisp		GKYPVHLSGGQQRVSIARALAMEPDLVLLFDEPTSAALDPELVG
malK		DRKPKALS GGQRVAIGRTLVAEP SVFLLDPELSNLDAAALRV
oppD		KMYPHEFSGGMRQVRMIAMALLCRPKLLIADEPTTALDVTVQA
oppF		NRYPHEFSGGCQRIGIARALILEPKLIICDDAVSALDVS IQA
RbsA	(N)	DKLVGDL SIGDQQMVEIAKVLSESKVIIMDEPTCALIDTETE
RbsA	(C)	EQAIGLLSSGNGQKVAIARGLMTRPKVLLILDEPTPGVDVGAKK
UvrA		GQSATTLSSGGEAQRVKLARELSKRGYILDEPTTGLHFADIQQ
NodI		NTRVADLSSGGMKRRRLTLAGALINDPQLLLILDEPTTGLDPHARH
FtsE		KNFPIQLSSGGEQQRVGIARAVVNKPAVLLADEPTGNLDDALSE

FIG. 16A

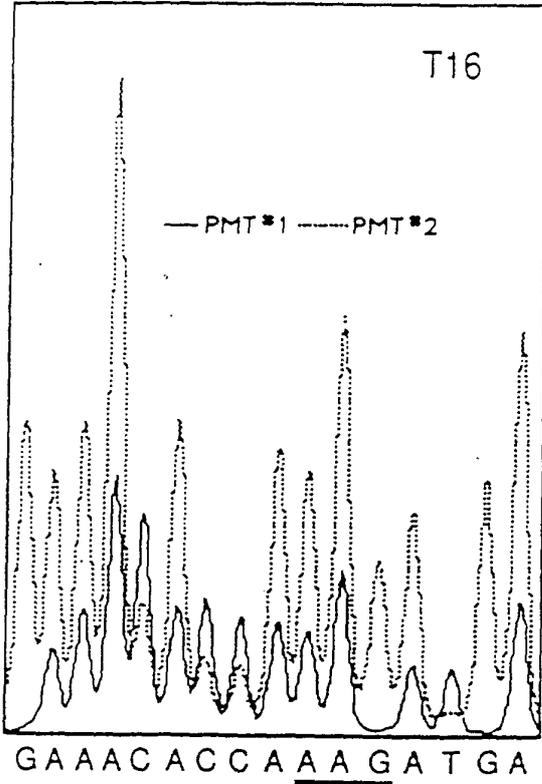


FIG. 16B

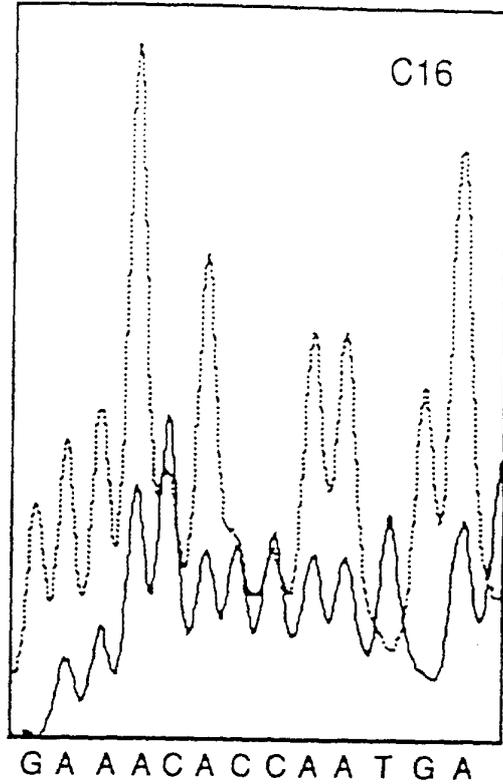


FIG. 17A

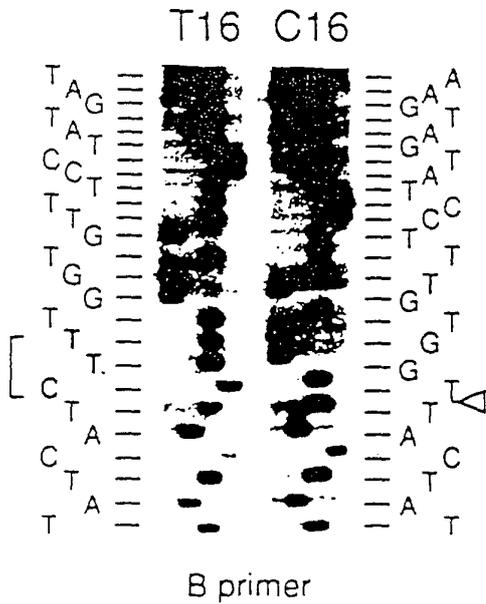


FIG. 17B

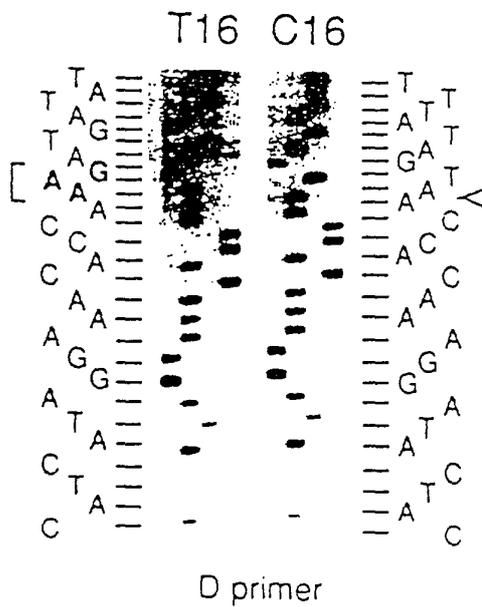


FIG. 18 (Fortsetzung)

AGGAATCTGCCAGNATATCGGCTGAGTGTGGTGTGTATGGTCTCCATGAGATTTTGICTCTATAATACTIGGGTAAATCTCCTIGGATATACTTGTG
.....
TGAATCAAACTATGTTAAGGGAAATAGGACAACACTAAATAATTTGGCACATGCAACTTATTTGGTCCCACTTTTATTCTTTTGCAG A GAA TGG GAT
Arg Glu Leu Ala Ser Lys Lys Asn Pro Lys Leu Ile Asn Ala Leu Arg Arg Cys Phe Phe Tip Arg Phe Met Phe
AGA GAG CTG GCT TCA AAG AAA AAT CCT AAA CTC ATT AAT GCC CTT CGG CGA TGT TTT TTC TGG AGA TTT ATG TTC
Tyr Gly Ile Phe Leu Tyr Leu Gly
TAT GGA ATC TTT TTA TAT TTA GGG GTAAGGNTCTCATTTTGTACATTTATGTATCACAATACTATNTGCATTTTGTGATTATGAAAAGA
CTAGGAAATCIGGGGAATAGGTGTAAAATAATATAAAGGATGATCCAACTCCAAACCACTAAGAACTCTAGTAAGGATAAGTAA.....
.....
CCACTATTCACTGTTTAACTTAAANTACCTCATATGTAAACTTGTCTCCCACTGTGGCTATAACAAATCCCAAGTCTTATTTCAAAGTACCAGATATTG
AAAATAGTGTCTAAGAGTTTACACATAIGGTAIGACCCCTCTATATAAACTCATTTTAAAGTCTCCTCTAAGATGAAAGTCTTGTGTTGAAATTTCTCAGGGT
.....
ATTTTATGAGAAATAAATGAAATTTTCTCTGTGTTTTTCCCTTTTGTAG GAA GTC ACC AAA GCA GTA CAG CCT CTC TTA CTG GGA
Arg Ile Ala Ser Tyr Asp Pro Asp Asn Lys Glu Arg Ser Ile Ala Ile Tyr Leu Gly Ile Gly Leu Cys
AGA ATC ATA GCT ICC TAT GAC CCG GAT AAC AAG GAG GAA CGC TCT ATC GCG ATT TAT CTA GGC ATA GGC TTA TTA TGC
Leu Leu Phe Ile Val Arg Thr Leu Leu His Pro Ala Ile Phe Gly Leu His Ile Gly Met Gln Met Arg
CTT CTC TTT ATT GTG AGG ACA CTG CTC CTA CAC GCC ATT TTT GGC CTT CAT CAC ATT GGA ATG CAG ATG AGA
Ile Ala Met Phe Ser Leu Ile Tyr Lys Lys
ATA GCT ATG TTT AGT TTG ATT TAT AAG AAG GTAATACTTCTTGGCACAGGCCCATGGCACATATATTCTGTATCGTACATGTTTTTAATG
TCATAAATTAGGTAGGAGGCTGGTACAAAGTAAAGGATAAATGCTGAAATTAATTTAATATGCTTAAATAAATGGCAGGATAAATTAATGCTCTTAAT
TATCCTTGATAAATTTAATTGACTTAACTGATAAATTTATTGAGTATC.....
TAAFTALICTGGCCTAGATGCGGGGAAATAAARCAACTAGAAAGCATGCCAGTAAATATGACGTGTGAAAGAAACATTTATGAACTGAGAAAGATAGTA
AGCTAGATGAATAGMATATAATTTTTCATTTACCTTTACTTAAATGAAATGCAATAAATGCTAATTAATTTTACTTATAATATATTGTA
.....
TTTTGTTTGTGAAATTAATCTAACTTTCCATTTTCTTTTAG ACT TTA AAG CTG TCA AGC CGT GTT CTA GAT AAA ATA AGT ATT
Gly Gln Leu Val Ser Leu Ser Asn Asn Leu Asn Lys Phe Asp Glu
GCA CAA CTT GAT GCT TCC AAC AAC CTG AAC AAA TTT GAT GAA GTATGTACCTATTGATTTAATCTTTTAGGCCACTATT
GTTATAAATTTATACAACCTGGAAAGGGGAGTTTTCTGGTCCAGATAAATAGTAAATAGTGGTAAAGTCTGCTCAGCTCAGCTCCTATTTCTGGAAAC
TAAAGAAAGGTCANTTGTATAGCAGAGCACCAATTTCTGGGCTGGTAGAACCCCAACTCAAGGCACCTTAGCCTGTTGTTAATAAGATTTTTCAAAAC
TTAAATCTTATCAGACCTTGTCTTTTAAAC.....

FIG. 18 (Fortsetzung)

.....
 GAATTCCACAAGGTACCAATTTAATTACTACAGAGTACTTTATAGAAATCAATTTAAATAATATAATAAATTGATAGATAGAGATTATATGCAATAAACAATTAA

 CAAAATGCTAATAAATACGAGACATATTGCAATAAAGTATTTATATAAANTTGATATTTATATGTTTTTATATCTTAMG C TGT GTC TGT AAA Cys Val Cys Lys Leu
 Met Ala Asn Lys Thr Arg Ile Leu Val Thr Ser Lys Met Glu His Leu Lys Lys Ala Asp Lys Ile Leu Ile Leu
 ATG GCT AAC AAA ACT AGG ATT TTG GTC ACT TCT AAA ATG GMA CAT TTA AAG AAA GCT GAC AAA ATA TTA ATT TTG
 His Glu Gly Ser Tyr Phe Thr Gly Tyr GGG ACA TTT TCA GAA CTC CMA AAT CTA CAG CCA TTT AGC TCA AAA CTC
 CAT GAA GGT AGC AGC TAT TTT TAT GGG ACA TTT TCA GAA CTC CMA AAT CTA CAG CCA TTT AGC TCA AAA CTC
 Met Gly Cys Asp Ser Phe Asp Gln Phe Ser Ala Glu Arg Arg Asn Ser Ile Leu Thr Glu Thr Leu His Arg Phe
 ATG GGA TGT GAT TCT TTC GAC CMA TTT AGT GCA GMA AGA AGA AAT TCA ATC CTA ACT GAG ACC TTA CAC CGT TTC
 Ser Leu Glu Gly Asp Ala Pro Val Ser Trp Thr Glu Thr Lys Lys Gln Ser Phe Lys Gln Thr Gly Glu Phe Gly
 TCA TTA GMA GGA GAT GCT CCT GTC TCC TGG ACA GMA AAA CMA TCT TTT AAA CAG ACT GGA GAG TTT GGG
 Glu Lys Arg Lys Asn Ser Ile Leu Asn Pro Ile Asn Ser Ile Arg Lys Phe Ser Ile Val Gln Lys Thr Pro Leu
 GAA AAA AGG AAG AAT TCT ATT CTC AAT CCA ATC AAC TCT ATA CGA AAA TTT TCC ATT GTG CMA AAG ACT CCC TTA
 Gln Met Asn Gly Ile Glu Asp Ser Asp Glu Pro Leu Glu Arg Arg Leu Ser Leu Val Pro Asp Ser Glu Gln
 CAA ATG AAT GGC ATC GAA GAG GAT TCT GAT GAG CCT TTA GAG AGA AGG CTG TCC TTA GTA CCA GAT TCT GAG CAG
 Gly Glu Ala Ile Leu Pro Arg Ile Ser Val Ile Ser Thr Gly Pro Thr Leu Gln Ala Arg Arg Arg Gln Ser Val
 GGA GAG GCG ATA CTG CCT CGC ATC AGC GTG ATC AGC ACT GGC CCC ACC CTT CAG GCA CGA AGG AGG CAG TCT GTC
 Leu Asn Leu Met Thr His Ser Val Asn Gln Gly Gln Asn Ile His Arg Lys Thr Thr Ala Ser Thr Arg Lys Val
 CTG AAC CTG ATG ACA CAC GGT AAC CAA GGT CAG AAC ACA ACA GCA TCC ACA CGA AAA GTG
 Ser Leu Ala Pro Gln Ala Asn Leu Thr Glu Leu Asp Ile Tyr Ser Arg Arg Leu Ser Gln Glu Thr Gly Leu Glu
 TCA CTG GCC CCT CAG GCA AAC TTG ACT GAA CTG GAT ATA TAT TCA AGA AGG TTA TCT CAA GAA ACT GGC TTG GAA
 Ile Ser Glu Glu Ile Asn Glu Glu Asp Leu Lys
 ATA AGT GMA GMA ATT AAC GMA GAA GAC TTA AAG GTAGGTATACATCGCTTGGGGGTATTTTACCCCCACAGAAATGCAATTGAGTAGAATG
 CAAATATGTAGCATGTACMAAANTTTACTAATAATCATAGGATAGGATAGGATATCTTAACACTCAGAAAGTATGAAGTTCATTTAATTATACAAAGCAAC
 GTTAAANTGTAAATAACMAATGATTTCTTTTTCGCAATGGACATATCTCTTCCCATAAATGGGAAGGATTTAGTTTTTGGTCTCTACTAAGCCAGTG
 ATAACCTGTGACTATAGTTAGAAAGCCATTTGCTTTATTACCATCTTGAACCCCTCTGTG.....

exon 13



FIG. 18 (Fortsetzung)

.....TTATTACTTATAGAAATAATAGTAGAAGACAAATAATGGTACCTACCCATTACCCACACACACCTCCAAATACCAGTAAACAATTTTAAAAAAGGGCAACACT
 TICCTAAATATTCAAATCGCTTTGGATTAAATCCCTGGTTGAATACCTACTATATGCGAGGCATTTATCTATTAGTAGAIGCIGIGATGACIGAGATTT
 AAAAATGTTAAATAGCCATAAATGGAAATGGTAAATTTAATGTGATATGTGCCCTAGGAGAAAGTGTGAATAAAGTCGTTCCACACAGAGAGAAATAAC
 ATGAGGTTCAATTTAGGTCCTTTTGTGCATCTATAG GA GAA GGA GAA AGA GTT GGT ATT ATC CTG ACT TTA GCC ATG AAT
 Ile Met Ser Thr Leu Gln Tip Ala Val Asn Ser Ile Asp Val Asp Ser Leu
 ATC ATG AGT ACA TTG CAG TGG GCT GTA AAC TCC AGC ATA GAT GTG GAT AGC TTG GTAAGTCTTATCATCTTTTAACTTTTA
 TGAAAAAATTCAGACAGTAACAAGTATGAGTAAATAGCAATGAGGAAGACTATATACCGTATATGAGCTTAAGNAATAAACAATTACAGATAAATTG
 AGGTCACCTGTATCTGTATCTTATCTCTTTCCCTTCATAGATAGCCACTATGAAGATCTAATACTGCAGTGAGCATTTCTTTACCCCTG
 TTTCCCTTATTCAGGATTTTCTAGGAGAAATACCTAGGGTGTGTTGCTGGTTCATAGGATTCACCCCATGCTTAAAC
TTCTCTTCAGTTAAACTTTTAAATATATCCAAATTTTCTCTGTAGTTCATTTGMAAAGCCCGACAAATAACCAAGTGACAAATAGCAAGTGTTCATTTT
 ACAAGTATTTTTTAGGAAGCATCAAACTAATTGTGAAATTTGCTGCCATTCITMAAACAAMAATGTTGTTATTTTATTTTCAG ATG CGA TCT
 Val Ser Arg Val Phe Lys Phe Ile Asp Met Pro Thr Glu Gly Lys Pro Thr Lys Ser Thr Lys Pro Tyr Lys Asn
 GTG AGC CGA GTC TTT AAG TTC ATT GAC ATG CCA ACA GAA GGT AAA CCT ACC AAG TCA ACC AAA CCA TAC AAG AAT
 Gly Gln Leu Ser Lys Val Met Ile Ile Glu Asn Ser His Val Lys Lys Asp Asp Ile Tip Pro Ser Gly Gly Gln
 GGC CAA CTC TCG AAA GTT ATG ATT GAG AAT TCA CAC GTG AAG AAA GAT GAC ATC TGG CCC TCA GGG GGC CAA
 Met Thr Val Lys Asp Leu Thr Ala Lys Tyr Thr Glu Gly Ala Ile Leu Glu Asn Ile Ser Phe Ser Ile
 ATG ACT GTC AAA GAT CTC ACA GCA AAA TAC ACA GAA GGT GGA AAT GCC ATA TTA GAG AAC ATT TCC TTC TCA ATA
 Ser Pro Gly Gln Arg
 AGT CCT GGC CAG AGG GTGAGATTTGAAACACTGCTTGTCTTTGTTAGACTGGTTCAGTAAAGTGAATCCCGTAGCCCTGAAGCAATGIGTATAGCAGA
 ATCTATTTGTAAACATTTATTTGTACAGTAGAATCAATATATTAACACACATGTTTTTATTTATATGGAGTCATTTTAAATATGAAATTTAATTTGCAGA
 GTCTGAACTATATAT.....

exon 19

Met Arg Ser

FIG. 18 (Fortsetzung)

.....
AAAGTCAGTGATAAAGGAAGTCTGCATCAGGGTCCAAATTCCTTATGGCCAGTTTCCTATTTCTCCAAGGTGGTTGTCTCCATATATCAACATTTG
GTCAGGATLGAAGGIGGCAACMAGGTTTGAATGATMGTGAANAATCTCCACTGGTGACAGGATAAATAATCCCAATGGTTTTTATTGAAGTACAATA
.....
CTGAATTATGTTTATGGCATGGTACCTATATGTACACAGAAGTGATCCCATCCTTTTACCTTATAG GTG GGC CTC TTG GGA AGA ACT CGA
Ser Gly Lys Ser Thr Leu Leu Ser Ala Phe Leu Arg Leu Leu Asn Thr Gly Glu Ile Gln Ile Asp Gly Val
TCA GGG AAG AGT ACT TTG TTA TCA GCT TTT TTG AGA CTA CTG AAC ACT GAA GGA GAA ATC CAG ATC GAT GGT GTG
Ser Trp Asp Ser Ile Thr Leu Gln Gln Trp Arg Lys Ala Phe Gly Val Ile Pro Gln
TCT TGG GAT TCA ATA ACT TTG CAA CAG TGG AGG AAA GCC TTT GGA GTG ATA CCA CAG GTGAGCAAAAGGACTTAGCCAGAA
AAAAGGCCAATAATTATATTTTTTACTGCTATTTGATACTTGTACTCAAGAAATTCATATTTACTCTGCMAAATAATATTTGTATTGCAATTCGTGCTTTT
TTTTCCAGGIGGAGLILICICAIAGGCAGAAAAGATGCTCTAAAAGTTTGGGAATTC.....
.....
TTTTTAATATTCTACAAATTAACAATTTATCTCAATTTCTTTTATTCTAAAGACATTTGGATTAGAAAATGTTICACAAGGGACTCCMAATATTCCTGTAGTAT
TTGTTTCTTAAAGAAATGATACAAAGCAGACATGATAAATAATTAAMAATTTGAGAGAACTTGTATGGTACATGGGTGTTTCTTATTTTAAATAATATT
.....
TTTCTACTTGAATAATTTTTACAATAACAAATAAGGGAAAATAAATAAAGTTATTTAAGTTATTCATACACTTTCTTCTTTCTTTTCTTTTGTCTATAG AAA
Val Phe Ile Phe Ser Gly Thr Phe Arg Lys Asn Leu Asp Pro Tyr Glu Gln Trp Ser Asp Gln Glu Ile Trp Lys
GTA TTT ATT TTT TCT GGA ACA TTT AGA AAA AAC TTG GAT CCC TAT GAA CAG TGG AGT GAT CAA GAA ATA TGG AAA
Val Ala Asp Glu
GTT GCA GAT GAG GTAAGGCTGCTAAGTGAATGATTTTGAAGGGGTAACATACCAACACAAATGGCTGATATAGCTGACATCATTCTACACAC
TTTTGTGTCATGTATGTGTGCACAACTTTAAATGGAGTACCCCTAACATACCIGGAGCAACAGGIACTTTTIGACTGGACCTACCCCTAACTGMAATGA
TTTTTGAAGAGGTAACATACCAACACAAATGGTTGATATGGTAAGATCATTTCTACACACTTTGTGTGCAATGATTTCTGTGCACAACTTCAAAAATGG
AGTACCCTAAAATACCTGGGGGACAAAGTACTTTTTCAGCTGAGCCTACTT.....

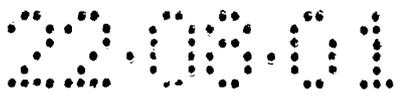


FIG. 18 (Fortsetzung)

.....CACAGTTGACTAATTTATGCTATCTTTTGGTCCCTCAGTCATGACAGAGTAGAAGATGGGAGGTAGCCACCCAGGATGNTGTCATACCTCCATCCCTTTATGCT
 ACATTTCTATCTTCTGTCTACATAAAGATGICATACACTAGAGGGCATACTGCAATGATACATATATCTTTCCAGCAGTGCATTCAGTTGTTGGAAATAA
 TTTATGTACACCTTTATMAAGCGTGGCCACAGGATGTCACAGCCATGTCGCCACGTATGTTTCTTACTACTTTTGGATACCTGGCCACGTAATAGACACTCATTG
 AAGTTTCCCTAATGAATGAGTACAAAGATAAACAACAGTTATAGACTGNTTCTTTTGGAGCTGTCAGGTTGTAATAGACTTTTGGCTCAANTCAATTCAAA
 TGGTGGCAGGTAGTGGGGTAGAGGGATTGGTATGMAAACAATAGCTTTCAGAACCTCCCTGTTTATTTTAGAATGTCAACTGCTTGGAGTGTTTTTAA
 exon 22 Val Gly Leu Arg Ser Val Ile Glu Gln Phe Pro Gly Lys Leu Asp Phe
 CTCTGGTATCTGAACATACTTCTCTMACTGCAG GTT GGG CTC AGA TCT GTG ATA GAA CAG TTT CCT GGG AAG CTT GAC TTT
 Val Leu Val Asp Gly Cys Val Leu Ser His Lys Gln Leu Met Cys Leu Ala Arg Ser Val Leu Ser
 GTC CTT GTG GAT GGG GGC TGT GTC CTA AGC CAT GGC CAC AAG CAG TTG ATG TGC TTT GCT AGA TCT GGT CTC AGT
 Lys Ala Lys Ile Leu Leu Asp Glu Pro Ser Ala His Leu Asp Pro Va
 AAG GCG AAG ATC TTG CTG CTT GAT GAA CCC AGT GCT CAT TTG GAT CCA GT GTGAGTTTCAGATGTTCTGTACTTAATAGCNC
 AGTGGGACAGMTCATTAAGCTTGGCTTCCAGACATATTTCTATTAGCCGTGTCATGCTGCGTGGGGTCTCCCAAGATATGNAATAAATTTGCC
 CAGTGGAAATGAGCATAAATGCATATTTCCCTTGCTAAGAGTTCTTGTGTTTTCTCCGAGATAGTTTT
 exon 23 1 Thr Tyr
 GCATGTTTATAGCCCAATMAAGAGTACTGGTGATTTCTACATAATGMAAATGTACTCATTTTAAAGTTTCTTTGMAATATTTTGTCCTGTTTATTT
 ATGGATACTTAGAGTCTACCCCATGGTTGMAAAGCTGATGGTAAAGACTTAAAGAAATAAAGTAATTTAAAGA
 GATAATAGAACAAATAGACATATTTATCAAGGTAATAACAGATCATTACTGTTCTGTGATATATATGTGTGGTATTTTCTTTTCTTAG A ACA TAC
 Gln Ile Ile Arg Arg Thr Leu Lys Gln Ala Phe Ala Asp Cys Thr Val Ile Leu Cys Glu His Arg Ile Glu Arg
 CAA ATA ATT AGA AGA ACT CTA AAA CAA GCA TTT GCT GAT TGC ACA GTA ATT CTC TGT GAA CAC AGG ATA GAA GCA
 Met Leu Glu Cys Gln Gln Phe Leu
 ATG CTG GAA TGC CAA TTT TTG GTGAGTCTTTATAACTTTTACTTAAGATCTCATTGCCCTTGTAAATTTCTTGATAACAATCTCACATGTA
 TAGTTCCCTGCMAATGGCAACAAATGTACAAAGTCTTTTCCMAAATATGTATCATAAGCCATCCAGCTTACTCAAAATAGCTGCACAAGTTTTTCACTT
 GATCTGAGCCCATGTTGGTGAATATAGTAATCTAAATGGCAGCATATTTACTAAGTTATGTTTATAAATAGGATATATAFACATTTTGAGGCGGTT
 TATTTGGACCCAAATCATAACATACTACTGTTTAAAGATTTTAAAAAGGTCCTGATTTCTTCAATAACTAATGTCCTCCATGGATGTGGGCTGG
 ACAGCCCTAGTTGCTTACAGTCTGATTTATGGTATTAATGACAAAGTTGAGAGGCCACATTTTCATTTTCTAGCCATGATTTGGGTTACAGGTAGTACGTT
 TCTCAACCCACTTCTCACTGTTCTTAAAAAACACTGTCACATGCCCCAGGACAGTGGCTTACATCTGTAATCCCAATACCTTGGGAGGCTGAGGGTGGG
 ATTACTTGGGCCCCAAGAAATTC.....

.....

FIG. 18 (Fortsetzung)

.....AGATGGTAGAACCTCCTTAGAGCCMAAGGACACAGCCAGUUAATGGTACATACCTGATTTGTTCAAAATGCCAAGGCTCTGGACATTCGATTCCTTTGACTTT
 TATTTCCCTTGAGCCCTGTGCCAGTTTCTGTCCCTGCTCTGGTCTGACCTGCTCCCTTCTGTCCAGATCTCACTAACAGCCATTTCCCTAG GTC ATA Val Ile
 Glu Asn Lys Val Arg Gln Tyr Asp Ser Ile Gln Lys Leu Leu Asn Glu Arg Ser Leu Phe Arg Gln Ala Ile
 GAA GAG AAC AAA GTG CCG CAG TAC GAT TCC ATC CAG AAA CTG CTC AAC GAG AGG AGC CTC TTC CCG CAA GCC ATC
 Ser Pro Ser Asp Arg Val Lys Leu Phe Pro His Arg Asn Ser Ser Lys Cys Lys Ser Lys Pro Gln Ile Ala Ala
 AGC CCC TCC GAC AGG GTG AAG CTC TTT CCC CAC CGG AAC TCA AGC AAG TGC AAG TCT AAG CCC CAG ATT GCT GCT
 Leu Lys Glu Thr Glu Glu Val Gln Asp Thr Arg Leu AH
 CTG AAA GAG GAG ACA GAA GAA GAG GTG CAA GAT ACA AGG CTT TAG AGAGCAGCATAAATGTTGACATGGGACATTTGCTCATGGA
 ATTTGAGCTCGTGGACAGTCACCTCAIGGMAITGGAGCTGCTGGMACAGTTACCTCTGCCCTCAGAAACAAGGATGAATTAAGTIIITTTTAAAMAANG
 AAACAATTTGGTAAGGGMAITGAGGACACTGATATGGTCTTGATATAATGGCTTCCGCAATAGTCAAAATTTGTGTAAAGGTACTTCAAAATCCTTGNAG
 ATTTACCACTTGTGTTTTGCAAGCCAGNITTTCCCTGMAAACCCTTGGCATGTGCTAGTAAATTTGGAAAGGCAGCTCTAAATGTCAATCAGCCCTAGTTGNTC
 AGCTTATTGCTAGTGNMACTCGTMAITTTGTAGTGTGGAGAGMACTGMAATCATACTTCTTAGGGTATGATTAANGTAATGATNACTGGMACTCAGC
 GGTTTATAAAGCTTGTATTCCTTCTCCTCCTCCCTGATGTTTAGAACACAACTATATTTGTTGCTAAGCANTCCAACTATCTCATTTCCANG
 CAAGTATTAGMAATACCACAGGMAACCACAGACTGCACATCAAAATATGCCCATTCACACATCTAGTGCAGTCAAGMAAGGACTCCAGATCCTGGA
 AATCAGGGTATGTTGTCAGGTCTACCAAAATCTCAATATTTCCAGATAATCACAAATACCTTACCTGGMAAGGCTGTTATATACTTTCCACAG
 GGGACAGGATGGTTCCTTACCTGGMAAGGGCTGTTATAATCTTTCCACAGGGACAGGATGGTTCCCTTGATGMAAGTGTGATAGCCCTTTCCCAAC
 TCCAGAAAGTGACAAAGCTCACAGACCTTTGMACTAGAGTTTAGCTGGMAAGTATGTTAGTGCMAATTTGTCACAGGACAGCCCTTCTTTCCACAGAAAGCT
 CCAGGTAGAGGGTGTGTAGTAGATAGGCCCTGGCCACTGGCCACTGTGGGTAGACACACATGAAAGTCCMAAGCANTTAGATGATAGGTTGATGGTGTATGTTTC
 AGGCTAGATGATGTACTICATGCTGTACACTAAGAGAAATGAGAGACACACTGAAAGAACCCAAATCATGAATTAGTTTTATATAGCTTCTGTGTTTTA
 TAAATTTGMAAGCAAAATTTTTCTAGGAAATATTTTTAATAATGTTTCAACATATATTACAAATGCTGATTTTAAAGAAATGATTAATGAAT
 ACAATTTGATMAAATAATTTTTATTTGMAATATTGACTTTTTATGGCACTAGTATTTTTATGMAATATATTGTTCCACAGCTGTATGATTTCCAGCCCA
 GGGTATATTAAACCAGGGCCATGMAATCACCCTTTGGTCTGGAGGGAGCCCTGGGGCTGATCGAGGTTGTTGCCACAGCTGTATGATTTCCAGCCCA
 CACAGCCCTTAGATGCAGTCTGMAAGAGATGGTACCACCACTGTGACTGTTTCCATCAAGGGTACACTGCCTTCTCMACTCCAAACTGACTCTTAAGA
 AGACTGCATTTATTTACTGTAAAGAAATATCACTTGTCAATAAADAATCCATATTTGTTGMAAGTGTGTTGTTTTCCAGATGCGGTTCACTTGTGAT
 GTTTCATCAGTCTCACTCCMATTTCATAGCTTTCATGGAAACATGMAACACCGAATCTGTCTTTTAGATATAGCCCTC.....

exon 24

FIG. 19

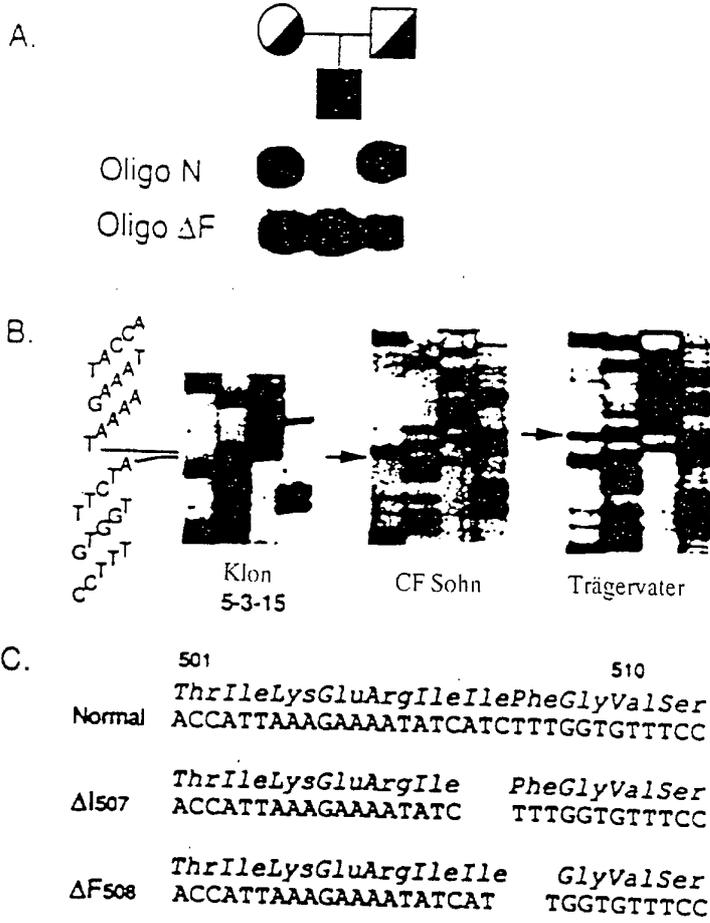
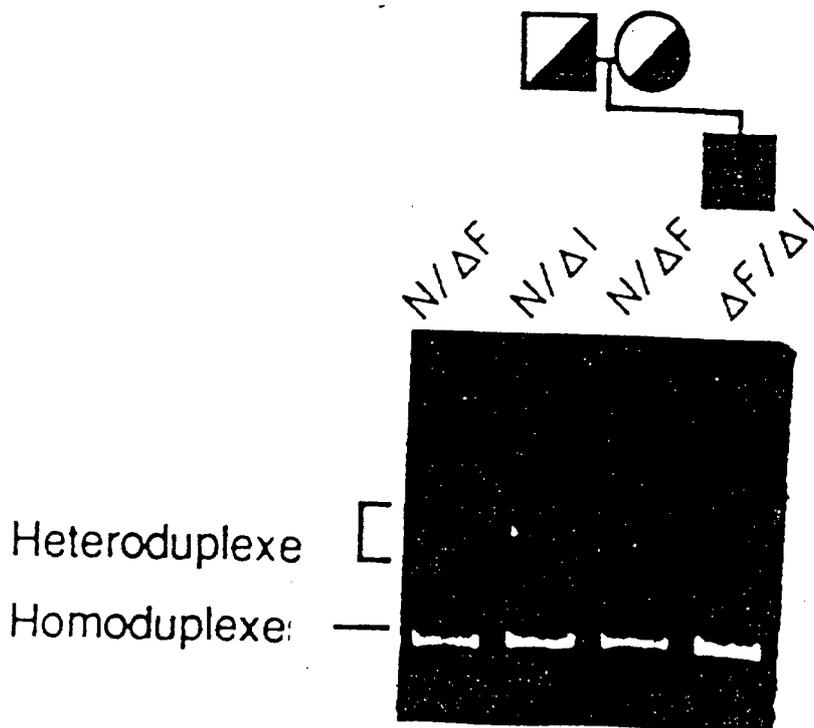


FIG. 20



2001

45/45

FIG. 21

