A 12 megabase restriction map at the cystic fibrosis locus

T.R.Fulton, A.M.Bowcock¹, D.R.Smith*, L.Daneshvar¹, P.Green, L.L.Cavalli-Sforza¹ and H.Donis-Keller

Department of Human Genetics, Collaborative Research Inc., Bedford, MA and ¹Department of Genetics, Stanford University, Stanford, CA, USA

Received August 26, 1988; Revised and Accepted November 21, 1988

ABSTRACT

We have constructed a physical map of the chromosomal region containing the cystic fibrosis locus using seven DNA markers and pulsed-field gel electrophoresis methods. The map includes cleavage sites for 8 rare-cutting restriction enzymes and spans over 12 megabases (Mb) of DNA, with one unlinked probe covering an additional 5 Mb. To our knowledge, this is the largest segment of human DNA which has been restriction mapped to date. We can identify thirteen putative HTF islands spaced at intervals of 0.3 - 3.2 Mb. The region between loci D7S8 and MET, where the CF gene lies, includes 1.4 - 1.9 Mb of DNA.

INTRODUCTION

Cystic fibrosis (CF) is a lethal genetic disorder which is inherited as an autosomal recessive at a frequency of about 1/2000 in North American and European Caucasian populations (1). It has been estimated that the frequency of CF heterozygotes may be as high as 1/20 persons (2). The major clinical manifestations of CF are chronic pulmonary disease and pancreatic enzyme insufficiency which appear to result from defective regulation of ion and fluid transport in certain epithelial cells (3,4). Since the primary biochemical defect has been difficult to study and still remains unidentified, one approach toward understanding the disease would be to first clone the gene and then identify the gene product.

The CF locus was mapped to chromosome 7 (7q31; ref.5) following the establishment of genetic linkage to a DNA marker, CRI-917 (D7S15; ref.6), which detects two distinct restriction fragment length polymorphisms (RFLPs). Other RFLP loci mapped by genetic linkage analysis to a 5 centiMorgan (cM) region surrounding the CF locus include: the MET oncogene (7,8), and the anonymous loci D7S8 (9), D7S18 (10), D7S23 (11), D7S73 (12), D7S97 (12), D7S99 (12), and D7S101 (12). Information from genetic crossovers in CF families has allowed CF to be placed between MET and D7S8 (13,14), with odds favoring this placement greater than 800:1 (14).

Since a number of these markers were likely to lie within 1 Mb of the CF gene and within a few megabases (Mb) of each other (assuming 1 cM = 1 Mb), it seemed feasible to use them to construct a long-range restriction map of the CF region by means of pulsed-field gel electrophoresis (PFGE) techniques (15-17). Such a map would establish the physical size of the region that is likely to contain the CF gene, and would resolve ambiguities which existed in the order of RFLP markers (12).

Several long-range restriction maps of human genomic regions have been constructed using PFGE methods. These include maps of up to 10 Mb at the Duchenne

Muscular Dystrophy locus (18-20), about 3 Mb at the HLA complex (21,22), two recent maps in the CF region covering 3-5 Mb (23,24), and shorter regions at other loci (25-28). Using seven of the markers mentioned above and eight restriction endonucleases that cut human DNA infrequently, we have used PFGE techniques to construct a restriction map spanning over 12 Mb of DNA at the CF locus.

MATERIALS AND METHODS

This work was carried out in two separate laboratories in which different methods were used. Most of the work in one lab used the lymphoblastoid cell line WT/51 as the source of DNA, and most of the work in the other was done using blood lymphocytes. In addition, the work took place over an extended period of time in which various methods were tried, and often modified or replaced. Therefore, in order to avoid confusion, the following sections are defined as follows: A. methods used in the Stanford lab with WT/51; or B. methods used in the CRI lab with blood lymphocytes. Only the preferred methods are given.

Preparation of High Molecular Weight Human Chromosomal DNA

A. Cells from the lymphoblastoid line WT/51 were washed several times at 4°C in diethylpyrocarbonate treated 10mM NaCl, 10mM Tris-Cl, 25mM EDTA, pH 7.5 (RSB; 29), and resuspended at a concentration of 30 million cells/ml in the same solution. Cells were warmed to 37°C, and an equal volume of 1% low melting agarose (Bio-Rad), 60 ug/ml proteinase K (in RSB) at 37-42°C was added. Aliquots of 100 µl were dispensed into sample holders and allowed to solidify. The inserts were placed in an equal volume of RSB with 1% SDS, and incubated overnight at 50°C. Inserts were then washed extensively at 50°C in 50 mM EDTA pH 7.5 and stored at 4°C.

B. Freshly drawn blood lymphocytes were prepared by a modification of the method of van Ommen and Verkerk (30). Cells were treated twice with isotonic lysis buffer as described, washed in 75 mM NaCl, 25 mM EDTA, pH 7.4 (SE) and resuspended in SE at 30 million cells per ml. The cell suspension (room temp) was mixed 1:1 with 1% low melting agarose (BioRad) in SE (50°C.) and cast into ribbons the width and thickness of a gel sample well. When solidified, these were cut into strips, resulting in inserts approximately 1 x 4 x 9mm in size. The inserts were suspended in 500 mM EDTA, 0.5% SDS, pH 7.5 (ES), 1mg/ml proteinase K and incubated overnight at 50°C. They were then washed in 4-5 changes of ES (50°C), equilibrated in 50 mM EDTA, pH 7.5 and stored at 4°C.

Restriction Digestion of High M.W. Human Chromosomal DNA

All enzymes were obtained from New England BioLabs (NEB) except SstII and some batches of NruI which were from Bethesda Research Laboratories; digestions were carried out under conditions recommended by the manufacturer. DNA samples in agarose inserts were: A. dialyzed against 3 changes of buffer (20 volumes) at 37°C for 15 minutes each, digested in 50µl with 20-25 units of enzyme overnight, transferred into 500 mM EDTA to stop the reaction, and loaded into sample wells; or B. equilibrated in 10 ml of restriction buffer at 4°C overnight, digested in 100µl of fresh buffer with 40 units of enzyme for 6h, transferred into 50 mM EDTA for 5 min, equilibrated in electrophoresis buffer overnight at 4°C and loaded into sample wells.

Pulsed-Field Gel Electrophoresis

Orthogonal field (OFAGE;15) or contour-clamped homogeneous electric field (CHEF;17) gels containing 0.7 - 1.5% agarose (either standard or low melting) and 0.25 - 1X TBE (100mM Tris, 100mM boric acid, 5mM EDTA) were run at 8 - 14°C, 50 - 250 volts and 10 sec - 50 min pulse times. In some cases, several pulse times were

used in a single run in order to expand the range of resolution; see the legend to Fig. 1 for an example. Molecular weight markers were: lambda phage DNA oligomer ladders (15) and chromosomal DNAs from the yeasts *Saccharomyces cerevisiae* (31; strains CGY339 and AB972), *Candida albicans* (32; strain ATCC 14053), and *Schizosaccharomyces pombe* (33; strain F642). *S. cerevisiae* strain CGY339 had identical chromosome sizes as the physically mapped strain AB972 (250, 280, 360, 450, 580, 580, 690, 780, 810, 835, 950, 970, 1125, 1125, 1600, and 3090 kb; A.Link, personal communication) except that the largest chromosome (XII) was slightly bigger.

Southern Transfer

Gels were stained in 1 µg/ml ethidium bromide, 1 mM EDTA. They were then: A. UV irradiated at 254 nm (1,500 microWatts per square cm for 2 minutes), denatured in 0.5 M Tris-Cl, 1.5 M NaCl, pH 7.5, blotted upside-down onto Genetran membranes in 10 X SSPE (1.8 M NaCl, 0.1 M sodium phosphate, 10 mM EDTA, pH 7.0), and baked for 30 min at 80°C under vacuum; or B. treated with 0.25 M HCl for 30 minutes, nicked and denatured in 0.5 M NaOH/1.5 M NaCl, transferred in the same solution, washed in 0.5 M NaOH/1.5 M NaCl (30 min), 0.5 M Tris-Cl pH 7.4/1.5 M NaCl (1 h), 10 X SSC (10 min), baked for 30 min at 80°C under vacuum, and washed for 1 h at 65°C in 0.1 X SSC/0.2% SDS (these washing steps were very effective in reducing background in subsequent probe hybridizations).

Probe labeling

The probes MET D&H (7,8), J3.11 (locus D7S8; 9), 7C22 (locus D7S18; 10), CRI-S14 (locus D7S73; 12), CRI-S158 (locus D7S97; 12), CRIS162 (locus D7S99; 12), and CRI-S167 (locus D7S101; 12) have been described previously. Purified insert fragments from these probes (or single copy subclones of the latter four; details available on request) were radiolabeled to a specific activity of 0.5 - 1.0 X 10⁹ dpm/µg: A. using random hexanucleotide primers (34) in agarose gel slices (35); or B. by nick translation after isolation on polyacrylamide gels (0.1 µg DNA, 0.2 mM each dATP, dGTP, and dTTP, 50 mM Tris-Cl, pH 7.5, 50 mM KCl, 5 mM MgCl₂, 5 mM mercapto-ethanol, 0.5 mg/ml BSA, 50 µCi α[³²P]dCTP (3000 Ci/mmol), 0.1 ng DNase I in 10 µl, incubated for 10 min at 37°C, cooled on ice, 5 units DNA polymerase I (NEB) added, incubated 7 min at 18°C, and stopped with excess EDTA).

Probe Hybridization

Filters were treated as follows: A. Blots were prehybridized for 6-18 hours at 42°C in 3 X SSPE, 50% formamide, 100 µg/ml sonicated salmon sperm DNA, 5 X Denhardt's, 1% SDS, the probe was denatured with 150 mM NaOH for 10 min, neutralized with HCl, added to the hybridization solution (as above plus 5% dextran sulfate) at a concentration of 4 million dpm/ml, and hybridized at 42°C for 2 days; or B. Blots were prehybridized for 1-16 h at 50°C in 1.5 M NaCl, 50 mM Tris-Cl, pH 7.5, 1 mM EDTA, 0.2% sarkosyl, 30% formamide, 10% dextran sulfate, 0.5 mg/ml heparin, 0.1% NaPPi, 0.1 mg/ml sonicated single stranded salmon DNA, they were then hybridized for 20 - 36 h at 50°C with 4 million dpm/ml of probe (denatured in 0.1 M NaOH, mixed with 5 ml of the above solution, and added to the prehybridization bag). Labeled yeast DNAs were included at 0.4 million dpm/ml to allow direct visualization of molecular weight markers on autoradiograms.

After hybridization the blots were washed in: A. 2 X SSPE, 0.2% SDS for 15 minutes at 42°C, in 0.5 X SSPE, 0.1% SDS for 15 minutes at room temp, and in the same solution at 60°C, for 2 X 30 minutes; or B. 3 X 30 min in 2X SSC at room temp, 2 X 30 min in 0.1X SSC/0.2% SDS at 60°C. All filters were blotted dry and exposed to XAR-5 film (Kodak) for 1 - 14 days with an intensifying screen at -80°C. Nylon filters

could be re-probed after stripping in 0.1 M NaOH for 30 min and neutralizing in 0.2M Tris-Cl, pH 7.5, 2X SSC, 0.5% SDS.

Multipoint Linkage Analysis

Multipoint linkage analysis was performed using a recently updated version of CRI-MAP (36, and P. Green, unpublished). Maximum likelihood analysis was done using genotype data on 23 reference families from the Centre d'Etude du Polymorphisme Humain (CEPH). The order of loci was fixed according to the physical mapping data determined in this study.

RESULTS AND DISCUSSION

Probes and Enzymes Chosen

We used seven probes that were known to map close to the CF gene and which were available to one or both of our labs. Genetic mapping indicated that these probes spanned about 5 cM flanking the CF gene (12).

Among all enzymes tested, the eight that produced the largest fragments were: *Mlu*I, *Not*I, *Nru*I, *Bss*HI, *Sst*II, *Nar*I, *Nae*I and *Sfi*I (shown in decreasing order of average fragment size generated in the CF region). These enzymes therefore offered the greatest possibility for generating fragments to which two or more probes would hybridize, thus enabling us to "link up" loci. The enzymes *Sst*II and *Sac*II recognize and cleave the same sequence, although *Sst*II was found to produce fewer partial digests. Both enzymes were used and will be referred to specifically in the text. However, to avoid confusion, only one designation is used on the maps in Figs.4 & 5 (*Sst*II).

Results of Southern Analysis

Each of the probes was first hybridized to Southern blots of either OFAGE or CHEF gels to determine the fragment lengths for each of the enzymes used. Field inversion gels were tested but were not used extensively because they tended to produce fuzzy bands and were limited in high molecular weight resolution. In most cases multiple gels had to be run under a variety of conditions in order to resolve all of the fragment lengths, since their sizes varied from 150 kb to 5Mb. When more than one probe appeared to detect the same sized fragment, this was confirmed by hybridizing the probes in question to the same filter. Southern blots of double digests were then hybridized to each probe using most or all of the 36 possible permutations of enzymes. The use of multiple switch-times during a single run was found to greatly expedite this process since in most cases both large and small fragments could then be resolved on a single gel. An example of a Southern blot from a gel run under such conditions is shown in Fig. 1. In this example fractionation occurred with a near linear relationship of fragment size to migration distance. In testing a variety of switch-time combinations and continuous ramps, we have never observed any anomalous migration of *S. cerevisiae* chromosomes. The use of low melting agarose (Sea Plaque; FMC) was observed to result in noticeably sharper bands and lower background; its fractionation properties were otherwise similar to conventional agarose.

Construction of Short Range Restriction Maps

For each probe, a short range restriction map was constructed. The construction of these maps was not trivial, so their derivation is briefly summarized below. The consensus maps are shown in Fig.4. It should be emphasized that during the course of this work a great deal of redundant data was collected by both of our groups, and the final maps were confirmed by numerous other double digests which are not described. Sometimes the maps contained ambiguities when there was more than one possible arrangement of sites that could account for the observed fragments. Most of

Fig.1 Autoradiograms of a CHEF gel Southern blot containing double digests of blood DNA against restriction endonuclease *NruI* probed with: A.CRI-pS158 (D7S97); B.CRI-pS167 (D7S101); C.CRI-pS162 (D7S99); D.J3.11 (D7S8); C=*C. albicans* molecular size markers (largest chromosome is about 3.2 Mb); S=*S. cerevisiae* chromosomes (size estimates for selected chromosomes are indicated in panel A). 1.no enzyme; 2.*NruI*; 3.*NruI/MluI*; 4.*NruI/NarI*; 5.*NruI/NotI*; 6.*NruI/SstII*. The gel (0.8% LMP agarose) was run in 0.5X TBE at 80 volts and 9°C for 5 days at a pulse time of 15 min, 1 day at 6 min, and 1 day at 80 seconds. On this particular blot, the *NarI* digest did not go to completion (lane 4). The smudgy band in lane 5 of the D7S101 panel is an artifact which did not appear on two other blots with *NotI/NruI* digests; this panel was also probed with a yeast *HIS3* clone to visualize *S. cerevisiae* chromosome XV.

these ambiguities were resolved when the short range maps were linked up, however small fragments that were not cleaved by any other enzymes could not be accurately placed without further information, such as partial digestions. When clusters of three or more sites were detected, these were presumed to be the locations for HTF islands (see below). Since we did not make extensive use of partial digest mapping, it is probable that some recognition sites for *SfiI*, *NaeI*, *BssHII*, and *SstII* were missed.

MET. The MET probes D and H detected the same fragments in all cases examined. These probes detected large *NotI*, *NruI* and *MluI* fragments (approximately 900 kb, 4Mb, and 5 Mb respectively) and small fragments with *BssHII*, *NaeI*, *NarI*, and *SacII* (250 - 350kb). The fragment sizes for *BssHII*, *NarI*, *NaeI*, and *SacII* did not change

appreciably in double digests involving various combinations of these four enzymes. This suggests that the cleavage sites for these enzymes are grouped together in two HTF islands flanking the MET locus and separated by about 300kb. Sfi I generated two fragments, 200kb and 570kb in size. This is due to an SfiI site that cuts slowly, which has also been observed by others (24,37). SfiI/NotI double digests generated fragments of 200 and 330kb, indicating that a NotI site lies about 130kb away from the slow-cutting SfiI site. NaeI/SfiI, NarI/SfiI, SacII/SfiI and BssHII/SfiI double digests all gave a 170-180kb fragment, unlike MluI and NotI which left the 200 kb SfiI fragment intact. MluI did not cut the 900kb NotI fragment so the MluI and NotI sites must map close together in one of the HTF islands.

D7S8 (J3.11). Most of the fragments detected by D7S8 were quite large: 1.2 Mb SfiI, 2.5 Mb SstII and BssHII, 3.5 Mb NarI, and approximately 4 Mb NruI, 4.3 Mb NotI (Fig. 2b), and 5 Mb MluI. Double digests of SstII, BssHII, and NarI against SfiI all produced a 1.0 Mb subfragment, suggesting the presence of an HTF island. NarI/NruI (blood DNA) gave a 2.6 Mb fragment (Fig.1d) of the same size as SstII and BssHII, suggesting the presence of another HTF island about 2.6 Mb away. D7S8 lies on a 350kb NaeI fragment. This was cut with SacII to yield a 290kb fragment. SfiI did not cut the NaeI fragment. Thus an NaeI site must lie 290 kb away from the SacII site and towards D7S99 (see below). The other site must lie 60kb away on the other side of the SacII site. A number of other small fragments were weakly detected by D7S8 (Fig. 1d), however these were determined to derive from one or more low level repeated sequences at other loci. See below for further discussion.

D7S99 (CRI-S162). D7S99 lies on a 2.2 Mb NruI fragment. This fragment was cut by MluI and NarI to produce subfragments of about 750kb, and by NotI to produce a fragment of 1 Mb (Fig. 1c). However, single digests with BssHII, NaeI, and MluI (in cell line WT/51) also produced fragments of 750kb. This places D7S99 at one end of the NruI fragment between two putative HTF islands, one of which contains an NruI site, and in WT/51 an MluI site. The other contains NarI, and MluI sites. We know from the long range map that SstII sites also occur within both HTF islands but the D7S99 SstII fragment is slightly smaller than the others (600 vs 750kb) so there is likely to be an extra SstII site near one of the islands. SfiI produced a 350kb fragment that was not cut by any of the other enzymes used (Fig. 4) and was not placed on the map. D7S99 detected some large fragments also: 3.5 Mb NarI, 4.3 Mb NotI, and 5 Mb MluI (seen in blood lymphocyte DNA only; see section on comparison with cell line DNA below).

D7S101 (CRI-S167). D7S101 lies within a 2.2 Mb NruI fragment. MluI/NruI double digests gave a 1.4 Mb fragment, and Not/NruI a 1 Mb fragment (Fig.1). The enzymes NaeI, BssHII, and NarI all produced fragments of about 300kb. This suggests the presence of two HTF islands flanking this marker which was confirmed in double digests amongst these enzymes and also by mapping other sites. SfiI/NotI, SstII/NotI and MluI/NotI gave fragments of 250-300kb although these enzymes all produced larger fragments individually (350kb SfiI, 700kb SstII, 4 Mb MluI, and 4.3 Mb NotI; Fig.2a). See Fig. 3 for examples of SfiI double digests.

D7S97 (CRI-S158). S158 hybridized to 2.2 Mb NruI, 4 Mb MluI, 1 Mb NotI, and 0.7 Mb BssHII, NaeI, SstII and NarI fragments. Again, the presence of several similarly sized fragments suggests flanking HTF islands. This was confirmed in double digests with various combinations of these enzymes, which resulted in the same size fragments. NruI/MluI digests produced a 1.4 Mb fragment (Fig. 1a), and as was the case with D7S99, a 350kb SfiI fragment was detected which was not placed on the map. The SstII and SfiI fragments observed at D7S97 were indistinguishable in size from

Fig.2 Autoradiograms of a CHEF gel Southern blot containing partial digests of blood DNA with restriction endonuclease NotI probed with: A. CRI-pS167 (D7S101); B. J3.11 (D7S8); C. MET H; D. 7C22 (D7S18). Molecular size markers (C & S) are the same as those described in Fig.1; approximate sizes of specific human fragments are indicated. 1.no enzyme; 2. 40 units NotI; 3. 20 units NotI; 4. 10 units NotI; 5. 2 units NotI; 6. 1 unit NotI. The gel (0.7% LMP agarose) was run in 0.5X TBE at 80 volts and 9°C for 2 days at a pulse time of 40 min, 2 days at 30 min, and 1 day at 20 min, 1 day at 6 min, and 1 day at 80 seconds. Size estimates for fragments less than 3 Mb were derived from other gels with better resolution in the appropriate size range.

those at D7S101, however in the latter case, both could be cleaved by NotI and, in the case of the SfiI fragment, MluI.

D7S18 (7C22). This locus lies within a 1.8 Mb NotI fragment. MluI and SstII gave 1.4 Mb fragments, NotI/MluI and NotI/SstII also produced 1.4 Mb fragments. NarI, NaeI and BssHII generated 600kb fragments. SfiI gave 400 kb and 570kb fragments, suggesting the presence of a slowly cutting SfiI site. The 570kb fragment was left intact by NotI, but was cut by MluI, SacII, BssH II, NarI, and NaeI to 425kb (Fig.3). These results suggest that the latter 5 sites reside in an HTF island close to an SfiI sites, and that two more islands lie 600kb and 1.4 kb away. This arrangement of sites was confirmed by

Fig.3 Autoradiogram of an OFAGE gel Southern blot containing double digests of cell line WT/51 DNA against Sfil probed with 7C22 (D7S18). Fragment lengths are indicated. 1.Sfil; 2.Sfil/BssHII; 3.Sfil/MluI; 4.Sfil/MluI; 5.Sfil/NarI; 6.Sfil/SstII; 7.Sfil/NotI. The gel (1.5% agarose) was run in 0.5X TBE at 250 volts and 14°C for 22 hours at a pulse time of 40 seconds.

partial Not I digests, which linked together the MET and D7S16 maps (see below and Fig.2).

D7S73 (CRI-S14). D7S73 hybridized to large fragments with most enzymes tested: 3.2 Mb NarI, 4 Mb NotI, 4.5 Mb MluI, and 3.5 Mb NruI, BssHII, and Sst II fragments. Double digests using all combinations of NotI, NruI, MluI, Nar I, and SstII gave fragments exceeding 3 Mb in all cases. This suggests that D7S73 is flanked by two HTF islands separated by at least this distance. The Sfil fragment is 670kb, and was not cleaved by any of the enzymes used in this study. Like J3.11, CRI-S14 cross hybridized weakly to one or more other loci (we have cloned one of these loci and determined that it derives from chromosome 4; DRS and TRF, unpublished results). This is particularly evident with enzymes SstII and MluI which revealed cross-hybridizing fragments of 1.6 Mb and 5 Mb respectively.

Link-ups Between Loci

D7S97, D7S99, and D7S101 all reside on an NruI fragment of the same size - - about 2.2 Mb (Fig.1). D7S101 and D7S97 both lie on a 1.4 Mb NruI/MluI fragment whereas D7S99 lies on an 0.75 Mb NruI/MluI fragment; these subfragments add up to 2.15 Mb. These results suggest that these probes may be linked together in the order: D7S97, D7S101, and D7S99. All other double digest data at these loci are consistent with this order, as well as the occurrence of D7S97 and D7S101 on a 4 Mb MluI fragment, and D7S101 and D7S99 on a NotI fragment of approximately 4.3 Mb.

In addition to D7S99 and D7S101, D7S8 also resides on a 4.3 Mb NotI fragment. This was confirmed to be the same fragment by the presence of an adjacent NotI fragment of approximately 1 Mb that was detected in partial digests (i.e. a NotI partial fragment somewhat larger than 5 Mb was observed; Fig. 2). This partial digest fragment was also weakly detected with MET. D7S8 was placed near the opposite end of the 4.3 Mb NotI fragment from D7S99 and D7S101 because it maps close to an HTF island containing a NarI site which occurs in that region (see above).

D7S8 detected several other large fragments, including a 3.5 Mb NarI fragment in common with D7S99, and a 5 Mb MluI fragment in common with both D7S99 (in blood lymphocyte DNA, see below) and MET. MET also hybridized to a 4 Mb NruI fragment in common with D7S8.

D7S18 lies on a 1.8 Mb NotI fragment. This is adjacent the 900kb MET NotI fragment, as indicated by the presence of a 2.7 Mb partial fragment, which was detected by both MET and D7S18 (Figs. 2c and 2d). An additional NotI partial with a size increment of approximately 500kb was also detected by D7S18. This fragment must lie on the opposite side of D7S18 from MET, since the map cannot accommodate any small NotI fragments on the D7S8 side of MET.

D7S73 hybridized to a 4 Mb NotI fragment, but the next fragment in partial digests was too large to be detected under the conditions used (>6 Mb). Also, unfortunately, there were no large fragments produced with any of the enzymes that allow this locus to be linked to the others. We know from genetic crossover data that D7S73 is closely linked to MET and is on the opposite side from D7S8, D7S97, D7S99, and D7S101 (12, 37). Therefore, considering the restriction mapping data, this locus must lie beyond D7S18.

This establishes the order: TER, D7S97, D7S101, D7S99, D7S8, MET, D7S18, D7S73, CEN.

The Long Range Restriction Map

Based on the locus order determined above, and the extensive mapping data on each probe, we were able to construct the long-range restriction map shown in Fig. 5a. In constructing this map it was necessary, as it is with any restriction map, to make adjustments in fragment sizes in order to piece together an internally consistent map. This usually involved slight to moderate reductions in the apparent sizes of the smaller fragments, which appear to be systematically overestimated on gels run at long pulse times and low voltage. Taking such adjustments into account, the long range map spans a contiguous region of about 12 Mb of DNA, with the (physically) unlinked probe D7S73 covering an additional 5 Mb. Since chromosome 7 represents about 5% of the genome (38), or about 170 Mb (5% of 3300 Mb), we estimate that the complete set of probes described here span approximately 10% of the chromosome.

Comparison with Genetic Linkage Data. We had previously constructed a genetic map of chromosome 7 which included all of these probes except D7S18. The order obtained (12) agrees with the physical order determined above, but there were insufficient informative meioses to uniquely localize loci D7S73, D7S99, and D7S101 with respect to D7S8 and MET. Re-analysis of the linkage data using the probe order determined above produced the map shown in Fig. 5b. According to this map, the most distal loci (D7S97 and D7S73) span a total distance of 3.5 cM and the distance between D7S8 and MET is 1.5 cM. Thus, the physical distance between D7S97 and D7S73 (>12 Mb) turns out to be much greater than expected on the basis of the accepted 1 cM = 1 Mb (on average) correlation, although the D7S8 - MET interval is close to the expected distance. On the other hand, loci D7S97 and D7S101 map very close to one another physically, but further apart genetically, with two recombination events occurring between them (Fig. 5). These loci are located in a region that contains many rare restriction sites and which also appears to contain some large scale repeated structure (see below for further discussion). This could possibly explain an elevated recombination rate.

Comparison with other CF physical mapping data. Two publications have recently appeared describing long-range restriction maps at the CF locus (23,24). The overall features of these maps are similar to ours, such as the presence of two HTF islands flanking MET at 240-300 kb, and the distances between loci D7S18, MET, and D7S8. However there are many restriction site differences between these maps and ours (and also between each other). These differences could result either from different

Fig.4 Short range restriction maps around individual loci. The heavy bar indicates the smallest fragment which was detected with each probe; the broken line indicates the full extent of the region in which the probes may lie. Fragments which could not be unambiguously placed without additional information are drawn either above or below the main line showing the probe location. Restriction sites not detected in blood DNA are marked by asterisks.

Fig.5 A. Long range restriction map. The position of each probe is indicated by a heavy bar. The restriction site abbreviations are as follows: B=BssHII, M=MluI, N=NaeI, R=NarI, T=NotI, U=NruI, F=SfiI, S=SstII; the scale is in Mb. Restriction sites not seen in blood DNA are marked by asterisks. **B. Multipoint genetic linkage map.** Genotype data from ref.12 was re-analyzed with the probe order fixed as determined in this study. The region containing the physically mapped loci is indicated by the bar; units are Kosambi centiMorgans.

patterns of methylation in the various cell lines used, from restriction site polymorphisms, or from errors in data interpretation. Without further comparative studies, it is not possible to conclude which of these is the major factor. We do wish to point out, however, two possible sources of error that other workers may have overlooked. First, it is very difficult to distinguish cross-hybridization of probes to secondary loci (as observed in this study with probe J3.11) from the presence of partial digestion products. Second, it is very important that appropriate electrophoresis conditions be chosen (and preferably more than one set of conditions) to ensure that adequate resolution is obtained and that fragment sizes are accurately assigned.

Comparison of Blood and Cell Line DNA. Two sources of high molecular weight DNA were used in this analysis. One was whole blood lymphocytes, the other was a lymphoblastoid cell line WT/51. In general, the restriction map generated from DNA of both sources was the same. However, an additional MluI site was detected near D7S99 in DNA from cell line WT/51 that was not observed in blood DNA. Thus, in this cell line the 5 Mb MluI fragment that hybridized to MET and D7S8 could not be detected with the D7S99 probe. Instead a 500kb MluI fragment was observed. The presence of this MluI site helped in constructing the short range map around D7S99, although its absence was useful in linking together the various loci. This underlines the advantage of using DNA from different sources in constructing long-range maps. Similarly, an MluI site was observed in the HTF island flanking MET that generated an MluI fragment of 300kb in cell line, but not blood DNA (Figs. 4&5). However in DNA from later passages of the cell line, this site was no longer present; possibly the result

of de novo methylation. This site appears on the maps of Poustka et al. (23) and Drumm et al. (24), both of whom used B-cell line DNA as source material. Within the D7S8 SfiI fragment of WT/51 DNA, an NruI site was observed that was not seen in blood DNA. This site was also described by Drumm et al. (24). In all, therefore, three extra restriction sites were observed in DNA from the cell line; it is not known whether these differences result from alterations in methylation or DNA sequence.

Where is the CF locus? CF has been mapped genetically between the loci D7S8 and MET (13,14). The minimum distance between these loci is approximately 1.4 Mb according to our restriction map (Fig.5a). Although many of the fragments spanning this region are quite large, the mapping data are all internally consistent, adding support to the size estimates given.

The NotI site between D7S8 and MET (Fig.5a) is in the same position relative to MET as one contained in an HTF island near a gene (IRP) which was once considered a candidate for the CF gene (11). RFLP probes at this locus show a high degree of genetic disequilibrium with CF (ie. specific alleles segregate together with the CF mutation a high proportion of the time), which is generally taken to indicate that the CF gene is close by (11). A CF diagnostic case which we studied, and 4 separate cases studied by others, revealed a recombination event between IRP and CF (37). Therefore we believe that the CF gene lies somewhere between the NotI site and D7S8.

Putative HTF Islands in the CF Region. Many vertebrate genes, and especially so called "housekeeping genes" which are required for general cell maintenance and are expressed in most cells, have regions known as HTF (HpaII tiny fragment) islands near their transcription start sites (40). HTF islands have a relatively high density of non-methylated CG dinucleotides (41) and consequently contain large numbers of HpaII and other, normally rare, CG containing restriction endonuclease cleavage sites (42). It has been estimated that 74% of SacII and BssHII sites, 42% of NaeI sites, and 89% of NotI sites in the human genome are localized in HTF islands (43).

Several putative HTF islands can be identified along the length of our restriction map by the presence of highly clustered recognition sites for several of the enzymes used (which, except for SfiI, contain one or two CG dinucleotides). For example, sites separated by approximately 300kb on either side of MET have such characteristics. This is not unexpected, given that MET is transcribed. HTF islands have been estimated to map 100kb apart on average in the mouse(41,42). In the region flanking the CF gene, we could identify thirteen putative HTF islands that were separated by an average of about 860kb (0.3 - 3.2 Mb). However, it is unlikely that all the HTF islands were located, especially if some are tightly clustered and/or do not contain recognition sites for any of the enzymes used in this study. One pattern which emerges from the map is that SfiI sites frequently flank HTF islands at moderate distances but only rarely occur within them.

Large Scale Repeats The region of DNA including the relatively closely spaced loci D7S97, D7S99, and D7S101 may contain a large-scale repeated structure. Probes for these loci all hybridize to remarkably similar sized SstII and SfiI fragments of about 700kb and 350kb respectively (those of S158 and S167 are indistinguishable except in double digests, see above). The probe J3.11 (D7S8) contains a sequence that cross-hybridizes to NarI/NruI, SstII, SstII/NarI, SstII/NruI, SstII/MluI, and SstII/NotI fragments that are indistinguishable in size from the corresponding D7S99 fragments (some of these can be seen in Fig.1). As part of a screening effort to obtain new probes in the CF region (to be reported elsewhere), we isolated a probe which is homologous to the 4 Mb D7S8 NruI fragment that also detects several fragments of

identical size as those mentioned above. This probe, however, detects unique fragments with other enzymes and maps to chromosome 4 by hybridization to a panel of hybrid cell lines containing various complements of human chromosomes. We suspect that most of these results can be explained by the presence of low level repeated sequences that occur on the order of 2 - 10 times in the genome. However, it is surprising that these sequences seem to occur on large restriction fragments of such similar size.

ACKNOWLEDGEMENTS

This work was supported by NIH grants 28428 to L.L.C.-S. and 37656 to H.D.-K.. We are grateful to A. Linn for performing cell culture, to D.Vollrath, J.Bell, D.Hardy, and D. Bowden for technical advice, to K. Falls for running linkage programs, and to R. Davis and G.Chu for CHEF gel apparati and plans. We also wish to thank S. Cartin-hour and E.Kauffman for subcloning CRI probes. B.Weiffenbach and T. Keith improved the manuscript with their comments.

*To whom correspondence should be addressed

REFERENCES

1. Talamo, R.C., Rosenstein, B.J., and Berminger, R.W., (1983) In J.B. Stanbury et.al. (eds), *The Metabolic Basis of Inherited Disease*, McGraw Hill, New York, pp.1889-1917.
2. Goodchild, M.C., and Dodge, J.A. (1985) In *Cystic Fibrosis*, Balliere Tindall, London.
3. Quinton, P.M. (1983) *Nature* 301, 421-422.
4. Li, M., McCann, J.D., Liedtke, C.M., Nairn, A.C., Greengard, P., and Welsh, M.J. (1988) *Nature* 331, 358-360.
5. Spence, M.A. and Tsui, L.-C. (1987) *Cytogenet. and Cell Genet.* 46, pp.170-187.
6. Tsui, L.-C., Buchwald, M., Barker, D., Braman, J.C., Knowlton, R., Schumm, J.W., Eiberg, H., Mohr, J., Kennedy, D., Plavsic, N., Zsiga, M., Markiewicz, D., Akots, G., Brown, V., Helms, C., Gravius, T., Parker, C., Rediker, K. and Donis-Keller, H. (1985) *Science* 230, 1054-1057.
7. Dean, M., Park, M., Le Beau, M.M., Robins, T.S., Diaz, M.O., Rowley, J.D., Blair, D.G. and Vande Woude, G.F. (1985) *Nature* 318, 385-388.
8. White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J.-M., Dean, M. and Vande Woude, G. (1985) *Nature* 318, 382-384.
9. Wainwright, B.J., Scambler, P.J., Schmidtke, J., Watson, E.A., Law, H.-Y., Farrall, M., Cooke, H.J., Eiberg, H. and Williamson, R. (1985) *Nature* 318, 384-385.
10. Scambler, P.J., Wainwright, B.J., Watson, E., Bates, G., Bell, G., Williamson, R. and Farrall, M. (1986) *Nucleic Acids Res.* 14, 1951-1956.
11. Estivill, X., Farrall, M., Scambler, P.J., Bell, G.M., Hawley, K.M.F., Lench, N.J., Bates, G., Kruyer, H.C., Frederick, P.A., Stanier, P., Watson, E.K., Williamson, R. and Wainwright, B.J. (1987) *Nature* 326, 840-845.
12. Barker, D. Green, P., Knowlton, R., Schumm, J., Lander, E., Oliphant, A., Willard, H., Akots, G., Brown, V., Gravius, T., Helms, C., Nelson, C., Parker, C., Rediker, Rising, M., K., Watt, D., Weiffenbach, B. and Donis-Keller, H. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 8006-8010.
13. Beaudet, A., Bowcock, A., Buchwald, M., Cavalli-Sforza, L., Farral, M., King, M.-C., Klinger, K., Lalouel, J.M., Lathrop, G., Naylor, S., Ott, J., Tsui, L.-C., Wainwright, B.,

- Watkins, P., White, R. and Williamson, R. (1986) *Am. J. Hum. Genet.* 39,681-693.
14. Lathrop, G.M., Farrall, M., O'Connell, P., Wainwright, B. Leppert, M., Nakamura, Y., Lench, N., Kruyer, H., Dean, M., Park, M., Vande Woude, G., Lalouel, J.-M., Williamson, and White, R. (1988) *Am. J. Hum. Genet.* 42, 38-44.
15. Schwartz, D.C. and Cantor, C.R. (1984) *Cell* 37, 67-75.
16. Carle, G.F. and Olson, M.V. (1984) *Nucleic Acids Res* 12, 5647-5664.
17. Chu, G., Vollrath, D. and Davis, R.W. (1987) *Science* 234, 1582-1585.
18. Van Ommen, G.-J.B., Verkerk, J.M.H., Hofker, M.H., Monaco, A.P., Kunkel, L.M., Ray, P., Worton, R., Wieringa, B., Bakker, E. and Pearson, P.L. (1986) *Cell* 47, 499-504.
19. Burmeister, M., and Lehrach, H. (1986) *Nature* 324, 582-585.
20. Burmeister, M. Monaco, A.P., Gillard, E.F., van Ommen, G.-J.B. Affara, N.A., Ferguson-Smith, M.A., Kunkel, L.M., and Lehrach, H. (1988) *Genomics* 2, 189-202.
21. Hardy, D.A., Bell, J.I., Long, E.O., Lindsten, T. and McDevitt, H.O. (1986) *Nature* 323, 453-455.
22. Lawrance, S.K., Smith, C.L., Srivastava, R., Cantor, C.R. and Weissman, S.M. (1987) *Science* 235, 1387-1380.
23. Poustka, A.M., Lehrach, H., Williamson, R., and Bates, G. (1988) *Genomics* 2, 337-345.
24. Drumm, M., Smith, C.L., Dean, M., Cole, J.L., Iannuzzi, M.C., Collins, F.S. (1988) *Genomics* 2, 346-354.
25. Woolf, T., Lai, E. Kronenberg, M. and Hood, L. (1988) *Nucleic Acid Res.* 16, 3863-3875.
26. Vollrath, D., Nathans, J., and Davis, R.W. (1988) *Science*, 240, 1669-1672.
27. Rappold, G.A. and Lehrach, H. (1988) *Nucleic Acids Res.* 16, 5361- 5377.
28. Gemmill, R.M., Coyle-Morris, J.F., McPeck, F.D., Ware-Vribe, L.F., Hecht, F. (1987) *Gene Anal. Techniques* 4, 119-131.
29. Ehrenberg, L., Fedorcsak, I. and Solymosy, F. (1974) *Prog. Nucleic Acid Res. Mol. Biol.* 16, 189-262.
30. Van Ommen, G.J.B., and Verkerk, J.M.H. (1986) In Davies, K.E. (ed), *Practical Approaches: Analysis of Human Genetic Diseases*, IRL Press, Oxford, U.K., pp. 113-133.
31. Carle, G.F., Frank, M., and Olson, M.V. (1986) *Science* 232, 65-68.
32. Vollrath, D., and Davis, R.W. (1987) *Nucleic Acids Res.* 15, 7865-7876.
33. Smith, C.L., Matsumoto, T., Niwa, O., Kico, S., Fan, J.-B., Yanagida, M., and Cantor, C.R. (1987) *Nucleic Acids Res.* 15, 4481-4489.
34. Feinberg, A.P. and Vogelstein, B. (1983) *Analyt. Biochem.* 132, 6-13.
35. Feinberg, A.P. and Vogelstein, B. (1984) *Analyt. Biochem.* 137, 266-267.
36. Lander, E.S., and Green, P. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2363-2367.
37. Julier, C and White, R. (1988) *Am. J. Hum. Genet.* 42, 45-48.
38. Ott, J. (1985) *Analysis of Human Genetic Linkage*, The Johns Hopkins University Press, Baltimore, Md. p.11.
39. Farrall, M., Wainwright, B.J., Feldman, G.L., Beaudet, A. Sretenovic, D., Halley, D., Simon, M., Dickerman, L., Devoto, M., Romeo, G., Kaplan, J.-C., Kitzis, A., and Williamson, R. (1988) *Am. J. Hum. Genet.* 43, 471-475.
40. Bird, A.P. (1986) *Nature* 321, 209-213.
41. Bird, A.P., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) *Cell* 40, 91-99.
42. Brown, W.R.A. and Bird, A.P. (1986) *Nature* 322, 477-481.
43. Lindsay, S. and Bird, A.P. (1987) *Nature* 327, 336-338.