Cystic Fibrosis: Recent Advances in Genetics and Molecular Biology

KENDALL KANE, M.D.

Good Samaritan Hospital and Health Center,
Dayton, Ohio 45406

ABSTRACT

Cystic fibrosis (CF) is the most common lethal hereditary disease of Caucasians, occurring once for every 2,000 live births. One out of 20 persons in the United States white population is a heterozygous carrier. An autosomal recessive pattern of inheritance is well established. The disease affects the respiratory and digestive tracts most severely. Despite the clearcut hereditary nature of the disease, insight into the biochemistry of CF has been almost totally lacking until very recently. New evidence strongly indicates that abnormal chloride ion transportation underlies the clinical manifestations.

Recent advances in molecular genetics have established that the CF disease gene is located on the long arm of chromosome 7. Several restriction fragment length polymorphisms (RFLP) markers are closely linked to the CF gene. These markers permit antenatal testing of samples from fetuses at risk for CF with a high probability of disease prediction. One laboratory has isolated a desoxyribonucleic acid (DNA) sequence from chromosome 7 which is a candidate for the CF gene itself.

A protein called the CF antigen is coded by a gene on chromosome 1. Patients with CF and carriers have abnormally high serum levels of this protein. In the normal state, the product of the CF gene on chromosome 7 may interact with the product of the gene on chromosome 1, enabling its normal catabolism and function. The structure of the CF antigen suggests that it may regulate ion transport.

Introduction

Cystic fibrosis is an inherited systemic disease which first manifests itself in childhood. Antibiotic therapy and modern pulmonary management have enabled many patients to live into adulthood. Death usually results from pulmonary insufficiency and right heart failure. Nutritional problems resulting from poor absorption of ingested calories and fat soluble vitamins often contribute to the clinical picture.

Reports of childhood steatorrhea and diarrhea can be found dating from the 1600’s. These may represent early descriptions of cystic fibrosis. According to folklore, a child who tasted salty had a poor outlook or was “bewitched.” Fanconi, Uehlinger, and Knauer are cred-
KANE

ited with the first modern case description in 1936. They noted the association of pulmonary and digestive symptoms and named the entity “cystic pancreatic fibromatosis and bronchiectasis.” At first, the disease was thought to be part of the celiac syndrome. Andersen and Hodges proposed in 1946 an autosomal recessive pattern of inheritance. Soon the disease was recognized as an entity separate from celiac disease. Andersen coined the term “cystic fibrosis.”

Genetics

Cystic fibrosis is the most common hereditary disease in the United States. Approximately five percent of Caucasians in the United States are heterozygote carriers. As a result, there is one case of CF for every 2,000 live births in this group.

Cystic fibrosis is much more common in some populations than in others. The highest incidence is in Central Europe and the British Isles. In one study from Scotland, one birth out of 500 was afflicted. Italians are rarely afflicted with a rate of one per 15,000 births. Blacks and Orientals have even lower rates. Orientals studied in Hawaii have an incidence of one per 90,000 births.

There is a strong selection pressure against the disease allele owing to the fact that homozygous afflicted individuals rarely reproduce. The continuing high frequency of the disease gene is unexplained and raises the possibility of an unknown selective advantage for the carrier state. Other possible explanations include continuing mutation; however, this is considered unlikely. No example of mutation has ever been observed. Cystic fibrosis is not associated with any structurally observable cytogenetic abnormality.

Physiology

A strict pattern of autosomal recessive inheritance points to a one gene-one enzyme disease. However, the exact nature of the causal biochemical defect has long resisted detection. Afflicted individuals demonstrate a consistent set of biochemical abnormalities. Ion transport abnormalities involving sodium, calcium, and chloride have been reported for decades. These have been considered to be the cause of the thickened bronchial and pancreatic secretions which are more directly responsible for the clinical symptoms. Increased salt concentration in the sweat of diseased individuals has been the basis of the “sweat test” which tests for increased sodium, chloride, or conductivity in the sweat. The sweat test is still the best available test for CF in clinical use at the present time, although its shortcomings are well recognized. Molecular genetics methods, which will be discussed, have been used to resolve equivocal sweat tests.

Recent studies indicate that abnormal chloride ion transport is fundamental to the CF disease process. Chloride ion must be actively transported in order to add water to bronchial and pancreatic secretions. A defect in this process causes abnormally thick tenacious secretions and obstruction of bronchioles and pancreatic ducts. Also, active reabsorption of chloride in sweat gland ducts is required to prevent excessive loss of salt in sweat. The sweat glands of CF patients show poor reabsorption of NaCl in the sweat duct.

Respiratory epithelial cells from CF patients fail to show activation of chloride ion transport channels on stimulation by epinephrine. However, normal chloride transport activity can be demonstrated in cell free membranes prepared from similar CF epithelial cells.
Adequate calcium ion must be present in the environment. Thus, while chloride transport is defective in whole cells from CF patients, the chloride transport channels in isolated membranes from CF patients are identical to those from controls. The fundamental defect in CF appears to lie in the regulation of chloride ion transport and not in the channels themselves. Frizzell et al. suggest that the problem lies between the intracellular cAMP mechanism and its ability to control chloride channel activity.

Cystic fibrosis patients have elevated levels of one or more serum compounds. Serum from CF patients causes ciliary dyskinesia in rabbit trachea and abnormal electrolyte transport in perfused rabbit salivary gland. A protein called the CF protein or the CF antigen has been found to be elevated in the serum of CF patients and heterozygotes. Quantitative immunoprecipitation and immunoradiometric assays have been developed for this antigen. Serum levels of the CF antigen can distinguish between normals, heterozygote carriers, and homozygous patients, who have the highest levels. However, the levels overlap somewhat. The test is technically difficult and antibodies have never been widely available; hence, the tests have not been widely used. The CF antigen is discussed further as a key to understanding the nature of the disease.

Efforts to Localize and to Identify the Cystic Fibrosis Gene

Tools

New techniques in the biochemistry of DNA are enabling rapid advances in our knowledge of the hereditary diseases. Researchers are using these tools to localize and to identify the site of the CF disease gene. Once the gene is identified, it will be possible to discover and to produce the protein which is encoded by this gene. It is hoped that this knowledge will provide the key to the pathogenesis of CF which has proven so elusive. It is also possible that the product of the normal allele will have direct therapeutic application. In any case, identification of the gene product will greatly enhance our knowledge of the CF disease process.

Restriction endonucleases (REns) are the principal biochemical tools for dissection of the genome. Derived from bacteria, they seek out specific base sequences several units long. These recognition sites are unique to each enzyme. Desoxyribonucleic acid is cleaved at these sites producing shorter DNA sequences called restriction fragments (RFs). Digestion of the entire genome, or any portion of it, results in a set of predictable RFs. These can be separated by electrophoresis into fragments of varying size. Repeated digestion by other RENs produces still smaller RFs.

The set of RFs which is produced from a genome or a subset of the genome is more or less constant within a species. However, individual variation within a species causes differences on the average of one per every 200 base pairs. Thus, the RF set which results from digestion with a given REN is a sort of biochemical fingerprint. Most of the genetic variation either is not expressed in the phenotype or is a benign variation. Obviously, some differences will be pathological. The mutation for Sickle cell disease occurs at a site which is cleaved by REN Mst II.

Variation in a DNA base pair is a polymorphism. If this occurs at a REN cleavage site, the result is alteration in the DNA digestion fragment lengths which are called restriction fragment length polymorphisms or RFLPs. For example, variation within an existing cleavage site will destroy its recognition by the REN. Hence, one longer fragment will replace...
two shorter ones. Conversely, a mutation may transform a sequence which had previously not been recognized by an REN into a sequence appropriate for cleavage. The result is that two short fragments will be seen following cleavage where previously there had been one longer fragment. Thus the RFLPs, or variation in the length of the REN cleavage fragments, are a direct expression of variation in the genome.

Most of the chromosomal DNA sequence does not act as functional genetic information. Therefore, most RFLPs have no known phenotypic consequences. They are said to be “anonymous.” During meiosis they are distributed with the chromosomes according to classical genetic principles. They function as artificial gene loci with alternative alleles which can be detected in the laboratory. The RFLPs can also be cloned and used as probes in exploring the genome.

**Localization**

Researchers have for years been seeking a genetic linkage to the CF disease gene. Large family pedigrees had been searched for coinheritance of the CF gene with other alleles. Polymorphisms, such as those used in paternity testing, have been used as markers. These include blood groups, enzyme polymorphisms, and, more recently, RFLPs. At first the results were entirely negative. By 1985, 40 percent of the genome had been excluded from possibility as the CF gene site.

The first linkage to the CF gene was to the polymorphic gene which controls production of the enzyme paraoxonase or PON. PON is an arylesterase found in human serum which hydrolyses paraoxon into p-nitrophenol and diethylphosphate. Linkage was found in a study involving 22 Danish and five English families, each with at least two children affected with CF. Coinheritance of the PON and CF alleles was statistically significant indicating that PON and CF loci lay on the same chromosome. However, they were not closely linked, and the chromosomal site of PON was unknown.

The CF gene was linked to other markers in rapid succession. Most of the new linkages were to anonymous RFLPs. The first of these was to a DNA sequence called D0CRI-917. The D0CRI-917 was also linked to PON, but the chromosomal site was still unknown.

In November 1985, a landmark issue of *Nature* contained a group of papers which established even closer linkages of the CF gene to several markers. The D0CRI-917 fragment, and therefore the CF gene, were proven to lay on chromosome 7. Another group found tight linkage between CF and the anonymous RFLP pJ3.11, which is known to lay on the long arm of chromosome 7. Two groups joined forces for a third demonstration of the chromosome 7 localization. One paper established strong linkage between CF and the oncogene MET, while another paper showed that MET is located on the long arm of chromosome 7. Other closely linked markers were established. Figure graphically represents the bracketing of the CF gene by the MET gene and pJ3.11.

A more recent issue of *Nature* illustrates two separate approaches to CF research. One paper by Estivill uses DNA probe techniques to examine more closely the chromosomal site of the CF gene on chromosome 7. The researchers used mouse-human hybrid cell lines with chromosomal contributions from both species. Isolated DNA fragments are used as probes to detect RFLPs after digestion by various RENs. Two probes, pXV-2c and pCS.7, have located RFLPs which are more closely linked to the CF gene than any previously known genetic marker. These markers have also been
Narrowing the Search for the CF Gene

Figure 1. The cystic fibrosis gene lies on the long arm of chromosome 7. It is flanked by the MET gene and by J3.11 gene, which is also called pJ3.11.6 (Reproduced with permission of the publisher.)

mapped to other established linkages to CF, such as the MET oncogene. Strong linkage disequilibrium was also shown for pXV-2c and pCS.7. One pXV-2c allele was associated with the CF gene 91.5 percent of the time. Further, the pCS.7 A2 allele was associated with the CF disease gene in 70 out of 71 chromosomes studied. The one exception occurred in a family in which CF also crossed over with MET and other markers. This family also included an apparent recombination in a previous generation. The author raises the question of sampling or laboratory error. In any case, strong linkage disequilibrium between CF and markers pXV.2c and pCS.7 is apparent.

Many functional genes, as opposed to non-functional DNA sequences, are associated with regions which contain clusters of non-methylated dinucleotide CpG.3 These clusters are susceptible to digestion by certain RENs and are called HTF islands, (for HpaII Tiny Fragments). The region of the CF gene has been searched, and an HTF island has been discovered in this region. This has raised hopes that the CF gene itself may have been located. However, the island must be considered to be a “candidate” for the CF gene until the sequence has been examined and specifically related to the disease process.

Another paper in this issue discusses the CF antigen.8 This protein has long been known to be elevated in the serum of CF patients and in heterozygote carriers.5 The CF antigen is the product of a gene on chromosome 1. It should be noted that there are no known mutations at the CF antigen site; rather, increases in the serum level follow from one or two CF disease alleles at the chromosome 7 locus whose product is unknown. Dorin et al.7 found that CF antigen was the product of normal and of leukemic granulocytes. Those authors purified the protein and performed partial amino acid sequencing. The next step was to use the amino acid sequence to prepare an oligonucleotide probe for the gene. Both the gene and its CF antigen product are now known in detail. The CF antigen has a molecular weight of 10,938 and is composed of 94 amino acids. The protein has strong homology with several mammalian calcium binding proteins. These are postulated to be major transducers of biological calcium signals. The role of the CF antigen may be to participate in regulation of chloride transport through a cyclic AMP mechanism. The unknown product of the normal allele at the chromosome 7 CF locus may interact with the CF antigen, perhaps converting it or combining with it to create a new compound necessary for proper chloride...
transport. Even a single dose of the chromosome 7 product suffices for normal function. This is demonstrated by the fact that CF carriers are phenotypically normal for all practical purposes. Thus, the isolation and identification of the chromosome 7 normal product is urgently important. The protein may have direct therapeutic usefulness.

**Antenatal Tests for Cystic Fibrosis**

When a couple has parented a CF child, subsequent pregnancies of that couple have a one in four chance of having the disease. The amniotic fluid of an afflicted fetus has abnormally low levels of various microvillar enzymes. Thus, measurement of amniotic fluid alkaline phosphatase has been used to predict whether or not a pregnancy of a high risk couple will produce a child with CF. The most sensitive test is an immunoassay using monoclonal antibodies against three major isoenzymes of alkaline phosphatase. The sensitivity of the test is 91 percent with a false positive rate of six percent over a series of 140 pregnancies of high risk couples.

Antenatal tests using second trimester amniotic fluid have been largely supplanted by techniques using RFLPs closely linked to the CF disease gene on chromosome 7. Most RFLP “alleles” do not exist in a firm linkage to the CF gene in a large family study, but linkage will probably be stable within the two or three generations under investigation. This is another way of saying that most RFLPs are in linkage equilibrium with the CF disease gene. This is a consequence of the fact that crossing over occurs between RFLPs and the CF gene in the population at large. Therefore, linkage analysis must begin with study of a living afflicted child; the linkage which is established is limited to the family at hand. Parents must be heterozygous for one or more genetic markers so that informative links can be followed as markers for inheritance of the CF gene. If only one parent is heterozygous for markers, then the mating is partially informative. About 82 percent of families are fully or partially informative for CF using MET and pJ.11 markers, which were the most closely linked until the report by Estivill. Most recently, a group of collaborators report that 97.5 percent of families are informative when the newest and most closely linked RFLP probes are used.

Samples for DNA studies may be obtained using chorionic villus sampling (CVS) as early as the ninth week of pregnancy. CVS is available only at large centers at the present time. False negative and false positive rates are estimated at two percent and six percent, respectively. Thus, predictions by RFLP analysis are more certain than those which use amniotic fluid enzymes. The advantages of early testing to the high risk couple are discussed by Pembrey et al. Amniotic fluid tests are still valuable for couples who do not possess informative DNA markers.

Farrall et al have reported success with carrier detection using markers pXV.2c, pCS.7 and others which are even more closely linked to the CF gene. Linkage disequilibrium has been so strongly established that certain marker haplotypes are very strongly associated with the CF disease gene. These haplotypes may serve as surrogates for the gene even in the absence of information from near relatives, e.g., in cases in which the CF-affected relative is deceased. The authors caution that such studies should be limited to populations in which the disequilibrium has been firmly established.

Spence et al have outlined the limitations of antenatal testing with RFLP genetic markers. Among the concerns are: (1) How tight is the linkage between the marker and the CF gene? The work-
markers? Markers with many alleles which are available in adequate percentages are informative for more couples; (3) Is there a living afflicted child to establish linkage? If an afflicted child dies in infancy, there is no way to proceed with the workup; and (4) Is it possible that some cases of CF may be due to mutations at loci other than closely linked to the chromosome 7 markers? Is there more than one disease allele at this locus? The CF cases which might be caused by such alleles would not obey the predictions made by using the known markers. All evidence thus far is compatible with the view that all CF cases are due to one allele. Thus, it is said that CF is "genetically homogeneous." The possibility of heterogeneity has not been finally excluded.15

An editorial by Harper16 draws some conclusions about recent CF genetic research. He notes that the region of chromosome 7 surrounding CF gene is genetically very stable, in contrast to the instability which surrounds the Duchenne muscular dystrophy locus. Harper points out that probes MET and pJ3.11 are very close to the CF gene with less than one percent crossover rates. Also, these probes are very informative. Data from families of various geographic and ethnic origins continue to support the view that CF is genetically homogenous.

The same methods of RFLP linkage analysis which are used for antenatal prediction of disease may be used within families of living CF patients to study near relatives. It is frequently possible to determine whether or not unaffected siblings, aunts and uncles, or cousins are carriers or are normal. Again, workup must begin with an afflicted relative. It is possible that the marker described by Estivill may be so tightly linked to CF that it may be used as an almost-exact CF gene surrogate. It is likely that the CF gene itself will soon be identified so that it can be probed directly, as is now the case the sickle cell disease.

Acknowledgments

The author wishes to acknowledge the assistance of Dr. Martha N. Franz who reviewed the manuscript prior to publication.

References


