Perspective

The Search for the Cystic Fibrosis Gene

RAY WHITE

PROGRESS IN MAPPING THE MAJOR GENETIC DISEASES CONtinues, with cystic fibrosis (CF) being the most recent one to yield to analysis by linkage in families (*1–3*). CF is the most common lethal genetic disease in Caucasian populations, but until recently it was known only by its phenotypic characteristics, such as heavy and viscous mucus in the lungs, pancreatic dysfunction, and an increased salt concentration in the sweat of affected individuals. A heavy toll in morbidity and mortality has been exacted; children with pulmonary distress survive on antibiotics and physical therapy, with intermittent periods of hospitalization. Although supportive therapies have improved over the past decade, the life expectancy of individuals with CF is still less than 30 years.

It has been accepted for some time that the increased salt concentration in the sweat of individuals affected with CF is likely due to a primary defect in chloride ion conductance; recent evidence indicates that the primary biochemical defect in the CF cell is probably not in the chloride channel itself but in its regulation (*4, 5*). The chromosomal location of the CF gene has been identified (*2, 3, 6*), and new molecular genetic technologies may make its cloning possible. For example, investigators might be able to find a DNA segment from the region that would identify a messenger RNA capable of correcting the defective regulation of the chloride ion channel.

Several events in the long search for the CF gene have occurred during the past year; knowledge of the location of the gene causing cystic fibrosis has progressed from "somewhere in the autosomal genome" to localization to a specific region of a specific chromosome. The precision of the localization may place the CF gene within a region as small as several hundred thousand base pairs, which might contain only 5 to 50 genes. Consequently, a number of investigators have begun intensive study of the region in order to identify the segment that is responsible for cystic fibrosis.

The chronicle of progress over the past year is not without interest. Frustration and concern had been mounting among investigators applying DNA markers in linkage approaches to CF. Although many markers had been tested, none had convincingly indicated linkage. The possibility that cystic fibrosis might not always be caused by lesions in the same gene, but rather might actually be a collection of diseases having a similar set of symptoms resulting from mutations at different genetic loci, was beginning to be taken seriously.

In August of 1985, however, at a meeting of gene mappers in Helsinki, a Copenhagen group reported that they had obtained significant evidence of linkage, not with a DNA marker but with a protein marker, the enzyme paraoxonase (PON) (*7*). This was good news; such evidence would only be obtained if in the great majority of individuals the disease was caused by mutations within the same gene. The chromosomal location of PON was not known, however, nor did its discovery seem imminent; it is a difficult enzyme to work with and is not readily expressed in hybrid cells. The linkage to PON, therefore, did not provide adequate information for further targeting of the CF locus.

The fortunate discovery of linkage to a DNA marker locus was the result of a collaboration between researchers in Toronto and Boston (*1*). Although this DNA marker was no closer to the CF gene than the PON locus, it did provide the important information that the CF locus is located on chromosome 7 (*6*).

The almost immediate subsequent discovery of two new marker loci, each apparently extremely close to the CF gene, emerged independently in each of two laboratories: from Salt Lake City, the MET locus (*2*), and from London, the locus defined by the probe J3.11 (*3*). These two tightly linked markers provided the next step toward a precise definition of the gene location. In the original sample sets, neither of these marker loci revealed any recombination with the CF locus, and each was estimated to be within a few percent recombination distance.

Confirmation of these initial results was quickly obtained. At a meeting in Toronto sponsored by the Cystic Fibrosis Foundation, a number of research groups from different laboratories agreed to participate in a joint effort to confirm the linkages between CF and the tightly linked markers.

The results of this collaboration have been compiled (*8*). In brief, over 200 CF-affected families, representing more than 1200 genotyped individuals, have now been studied with the markers. The original findings are amply confirmed; a few recombinants have been obtained, but it is now clear that the recombination distances between each of the two closely linked markers, MET and J3.11, and the CF locus are almost certainly less than 1 percent.

The findings on gene order are still somewhat equivocal. The most likely order is MET-CF-J3.11. However, the next most likely order, CF-MET-J3.11, is only ten times less likely. The order CF-J3.11-MET is more strongly excluded, with the odds against it being 160 to 1. The question of gene order is important in that it bears on both diagnostic interpretations and on various schemes to identify the gene. However, the issue may soon be resolved as the recombinant chromosomes become better characterized for the marker loci. Additional closely linked marker loci have now been identified by both the London (*9*) and Boston (*10*) groups, but neither the order of these markers nor their distances from the CF locus relative to the earlier markers have been determined.

The available data support the notion that cystic fibrosis is almost always caused by a defect in the same gene. It is still possible, of course, that in a small number of affected individuals the disease could be due to some other gene. The most important aspect of the recent findings, however, is the unexpectedly high degree of precision of the localization. The chance finding of several marker loci so closely linked to the CF gene may suggest that recombination is relatively infrequent in this region of chromosome 7, and that the markers are actually widely spaced physically.

It may not be easy to isolate the CF gene and characterize its defect. If any mutant alleles of CF are the result of a deletion of DNA in the gene, it may be possible to use the deletion end points, as has been done for the Duchenne muscular dystrophy locus (*11*), to narrow the search to an even smaller region of chromosome 7. However, it is possible that this disease, like sickle cell anemia, may result from a single mutation or a very small number of mutations

The author is an investigator at the Howard Hughes Medical Institute, Research Laboratories, University of Utah, 701 Wintrobe Building, Room 631-B, Salt Lake City, UT 84132.

that may not be deletions. Furthermore, for a recessive disorder as prevalent as CF, the high frequency of mutant alleles in the population makes the detection of new mutations, which would perhaps include deletions, quite difficult.

One of the immediate goals of CF research is to identify the genes in the region of the CF locus. A number of DNA segments within the region defined by the markers must be identified, and a means of testing those DNA segments for the presence of the CF gene must be devised. Both requirements bring us into areas of new technology. It is now possible, for instance, to construct large-scale restriction site maps of regions of mammalian genomes covering thousands of kilobases, with the use of pulsed-field gel electrophoresis to resolve the large restriction fragments (*12, 13*). Large-scale "chromosome walking" technologies are also feasible (*14, 15*). Furthermore, because the linked genetic markers are defined by DNA probes, it is also possible to relate the physical map of restriction sites to the linkage map of markers linked to CF, in such a way as to bracket a physical region of DNA that must contain the CF gene. At this point, the path to the CF gene may take any of several branches, but in general each is likely to involve the use of cloned DNA segments from the physically defined region to identify or select genes or messenger RNA's that can correct the CF defect in some functional test.

Such a functional test will probably have to take into account recent evidence that the primary defect seems to be in the regulation of the chloride channel rather than in the channel itself. Experiments now indicate that if membrane patches are excised from the cell, chloride channel activity appears whether the patch is from a CF cell or from a normal cell (*4, 5*). These results do not eliminate the possibility that the defect could be in a regulatory domain of the chloride channel protein. However, they do suggest that investigators must take seriously the possibility that the primary defect is in a separate regulatory protein, one not encoded by the chloride channel gene. The consequences of a defect in a regulatory protein could be heterogeneous from one cell type to another; this might account for the wide range of phenotypic effects seen on different tissues of a CF-affected individual.

The implications of the linkage findings for clinical diagnostic purposes are profound. Genotypic diagnosis of an affected fetus can be made as early as 11 to 12 weeks into pregnancy through chorionic villus sampling, whereas a diagnostic determination based on enzyme levels in amniotic fluid cannot be made reliably before the 14th to 15th week of gestation (*16, 17*). The earlier diagnostic window can be important. The DNA method will also permit definition of carrier status among siblings and other close relatives of affected individuals.

The current limiting feature in DNA diagnosis is the informativeness of the marker systems. With the two well-characterized marker loci MET and J3.11, 50 to 80 percent of pregnancies are fully informative with respect to CF status. Many of the remainder are informative to the extent that sometimes CF can be ruled out. However, a number of completely uninformative or partially informative pregnancies are left with intolerable ambiguities (*18*). There is reason to be optimistic that the picture will soon improve; established methodologies provide for expansion of the informativeness of the marker loci, and the additional markers mentioned above should further improve the applicability and accuracy of diagnosis.

Also significant for diagnostic purposes is the question of gene order. Although each of the established markers is within a 1 percent recombination distance of the CF gene, it would be reassuring to know for each specific case whether a recombination event had occurred between the marker and the CF gene. If the marker loci flank the CF gene, a recombinant between a marker and CF will be detected as a recombinant between the marker loci. In such a case, the diagnostic information is unreliable because it would not be known which of the two intervals was recombinant. Although the current best estimate places the two established marker loci (MET and J3.11) as flanking the CF gene, the support for that order is not yet solid. However, as the current loci are made more informative, and the additional markers become integrated in the map, it will be possible to determine the order with increasing confidence.

Improvements in diagnosis, however, are not the only benefit hoped for in the search for the CF gene. When the gene is finally isolated and characterized it may well be possible to imagine more direct therapeutic interventions than those currently available.

REFERENCES

1. L. C. Tsui *et al.*, *Science* **230**, 1054 (1985).
2. R. White *et al.*, *Nature (London)* **318**, 382 (1985).
3. B. J. Wainwright *et al.*, *ibid.*, p. 384.
4. R. A. Frizzell, G. Rechkemmer, R. L. Shoemaker, *Science* **233**, 558 (1986).
5. M. J. Welsh and C. M. Liedtke, *Nature (London)* **322**, 467 (1986).
6. R. G. Knowlton *et al.*, *ibid.* **318**, 380 (1985).
7. H. Eiberg *et al.*, *Cytogenet. Cell Genet.* **40**, 623 (1985).
8. A. Beaudet *et al.*, *Am. J. Hum. Genet.*, in press.
9. P. J. Scambler *et al.*, *Nucleic Acids Res.* **14**, 1951 (1986).
10. H. Donis-Keller *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
11. L. M. Kunkel *et al.*, *Nature (London)* **322**, 73 (1986).
12. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
13. G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986).
14. F. S. Collins and S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6812 (1984).
15. A. Poustka and H. Lehrach, *Trends Genet.* **2**, 174 (1986).
16. R. Williamson *et al.*, *Lancet* **1981-II**, 1125 (1981).
17. D. Brock *et al.*, *Prenatal Diag.* **5**, 93 (1985).
18. M. Farrall *et al.*, *Lancet* **1986-I**, 1402 (1986).

25 August 1986; accepted 2 October 1986