Chapter 4

New Technologies for Detecting Heritable Mutations

Contents

TABLE

FIGURES

Chapter 4

New Technologies for Detecting Heritable Mutations

INTRODUCTION

New developments in molecular biology have opened the way for direct examination of large portions of human DNA to detect kinds and rates of mutations. Until recently, it has been possible only to infer the occurrence of genetic damage from indirect, limited, and often imprecise observations (see ch. 3). In principle, these new technologies would allow samples of DNA from people exposed to potential mutagens to be examined for precise genetic damage.

Little is known about the nature and frequency in humans of the majority of mutational events, since previous information was available only on DNA sequences that could be selected on the basis of gene expression. Until recently, human mutation was studied in fine detail only in particular genes (e.g., hprt and globin genes) that could be fished out of genomic DNA by using some gene-specific probe, or in genes whose products (mRNA, proteins, etc.) could be isolated and examined.

In this chapter, new methods are described for examining proteins and for examining large regions of genomic DNA (see table 6), regardless of their function or location in the genome and without devising selection methods for specific known genes. These new technologies are currently under development or being considered by the scientific community. Most of them were explored as potentially feasible technologies at a workshop cosponsored by the International Commission for Protection Against Environmental Mutagens and Carcinogens (ICPEMC) and the U.S. Department of Energy, in December 1984.

Table 6.—Current and New Methods for Detecting Human Heritable Mutations

Indicators of mutation	Method of detection
Current methods:	
Sporadic genetic disease	Sentinel phenotypes
Numerical or structural chromosome abnormal i ties	Cytogenetic analysis
Variant blood proteins	1 D-electrophoresis Quantitative enzyme analysis 2D-electrophoresis
New ***methods:***	
Altered DNA sequences	DNA sequencing Restriction fragment length polymorphisms 1 D-denaturing gel electrophoresis 2D-denaturing gel electrophoresis Ribonuclease cleavage Subtractive hybridization Pulsed field gel electrophoresis

SOURCE Office of Technology Assessment

Several different approaches are used in these new technologies. Two-dimensional electrophoresis of proteins follows from the techniques of one-dimensional electrophoresis and incorporates new methods of analyzing protein spots on the gel. In several of the DNA methods, restriction enzymes are used to fragment genomic DNA before sizing and sequence analyses are performed. Other technologies hinge on the production of hybrid DNA molecules, or heteroduplexes, and on gradient denaturing gels that are used to identify heteroduplexes with base pair changes. Another method devises a way of separating larger fragments of DNA than were previously possible to separate. These new methods are discussed below,

SHORT DESCRIPTIONS OF THE NEW TECHNOLOGIES

Two-Dimensional Polyacrylamide Gel Electrophoresis

A development following from the experience with one-dimensional electrophoresis for protein variants (see ch. 3) is two-dimensional polyacrylamide gel electrophoresis, "2-D PAGE, " the only technique described in this chapter that is currently in use in studies of human beings. With this technique, proteins are separated first by isoelectric focusing on the basis of their molecular charge, and then by electrophoresis on the basis of their molecular weight. From 500 to 1,000 proteins may be visible from a single sample on one gel, and about 100 of these may be sufficiently clear to detect rare variants in a child that are not present in the parents (92).

To search for new mutations in proteins using 2-D PAGE, blood samples are taken from child/mother/father trios. The program that has had most experience with this technique for mutation research, the University of Michigan Medical School, collects cord blood from newborns and parental blood samples at the time of the child's birth. Each blood sample is fractionated into six parts: platelets, plasma, nonpolymorphonuclear cells, polymorphonuclear cells, erythrocyte (red blood cell) membranes, and erythrocyte cytosol (the contents of the cell) (95). Each fraction is treated and analyzed separately.

Separation in the first dimension is by the charge of the protein molecules, qualitatively similar to the separation using one-dimensional electrophoresis. Each sample is run on a long, thin (1 to 2 mm in diameter) cylinder of polyacrylamide in a hollow glass tube using a technique called isoelectric focusing. The separation time can be speeded up or slowed down depending on the voltage applied. After the first separation, the gel, which looks like a clear noodle, is "extruded" from the tube.

In the second separation, the "noodle" is laid across the top of a gel about 7 inches square, that is held between two glass or plastic plates. A current is passed through the gel, and the protein molecules migrate down through the gel at a speed that is proportional to their size. A final step renders the proteins visible by autoradiography (for proteins that can be isotonically labeled) or by stains such as Coomassie Brilliant Blue and silver- or nickel-based stains.

The trained eye or a computer program detects patterns of spots on the gel, though most of the stained areas are too crowded or indistinct for drawing conclusions about specific proteins. Only spots in relatively clear areas and of a sufficient intensity are chosen for scoring. The analysis entails comparing the parent/child trios for differences. No two gels are run in precisely the same way, so it is often difficult to superimpose a set of three gels over one another and to look for differences. While protein spots bear the same relationships to each other on different gels, they routinely appear at slightly different coordinates.

For each type of sample mentioned above, a standard constellation of protein spots is scored for genetic variation. A mutant protein can appear on the gel as a spot for which there is no corresponding spot on parental gels. The gels can be interpreted visually, which is the traditional method, and currently the most accurate, but computer-assisted interpretation is desirable. Two steps in analysis lend themselves to computerization. First, and now routinely done, the protein spots can be "read" by computer for location and intensity. The second step is computerized comparison of different gels. Two approaches for comparison are being developed in different laboratories, each of which will probably be useful on a "production scale" in the near future (92). Eventually, computers will be relied on to screen gels, and to eliminate from further consideration the overwhelming majority of parent/child trios with no variants.

In comparison to one-dimensional protein separation, a relatively small number of parent/child trios have been analyzed with 2-D PAGE, and no heritable mutations have been found. However, the results of these analyses have revealed a wide array of largely unknown proteins, and distinct differences in the spectrum of proteins from different cell types. The range of electrophoretic vari-

ants of a number of these proteins have been characterized in plasma (104) and in erythrocyte lysate (105). In plasma, the variation is substantially higher than has been reported in other blood fractions.

The advantage of 2-D PAGE is that it can be done now without a major developmental breakthrough and without a great initial cost, though the running expenses are significant. It may be some years before any of the DNA analytic methods catch up to protein analysis in practical application. The future of protein analysis for detecting mutations after DNA methods become routine is unclear, however.

There are technical limitations of the 2-D PAGE technique. First, the technique itself is technically demanding. Second, the number of proteins that can be scored on each gel is limited by the degree of separation possible on plates that are of a practical size.

As with most of the other techniques, only a certain spectrum of mutations is detectable with 2-D PAGE. Chromosome rearrangements in general are not revealed through electrophoretic variants. This technique has its greatest strength in detecting point mutations and larger deletions in the DNA that change the size or charge of a protein molecule. Point mutations that do not cause such changes, and small insertions or deletions, are less likely to be detectable with this method. Currently, null mutations cannot be detected reliably, and this is probably the greatest disadvantage of 2-D PAGE. Overall, 2-D PAGE should detect an estimated 25 to 35 percent of all spontaneous gene mutations. Several aspects of 2-D PAGE of blood proteins, in addition to computerized comparisons of gels, are amenable to improvement, and at least one approach is being developed to improve detection of null mutations (92).

The identity of most of the proteins that can be visualized on two-dimensional gels is unknown, Although it has not yet been done, the technology exists to recover proteins from gels, purify them, and then determine their amino acid sequences (92). Once amino acid sequences are worked out, the nucleotide sequences that code for them can be deduced and nucleic acid probes can be assembled to correspond to the protein-coding nucleotide sequences. The probes could theoretically be used to locate the corresponding DNA within the human genome, thereby determining exactly where in the genome the coding sequence lies.

DNA Sequencing

Conceptually, the most thorough and straightforward analysis of the genome would consist of lining up the chromosomes and determining the nucleotide sequence from end to end on each one. Once the sequence was known, it could be stored in computer memory.

The advantage of determining the complete genomic sequence would be that any nucleotide sequence that was identified in any laboratory could immediately be mapped to its chromosomal location. As nucleotide sequences for various parts of the genome were obtained in different laboratories, both common polymorphisms and rare mutations could be identified by making comparisons between the total genomic sequence and partial sequences. Furthermore, the method would detect mutations and polymorphisms no n-tatter where they occur—in introns, exons, and repetitive sequences.

Walter Gilbert, Fred Sanger, and their colleagues developed the theoretical basis for the techniques that makes it possible to consider sequencing the genome (67,118) and in 1984, George Church and Walter Gilbert described a method for sequencing any part of the genome (2o), although the method could be applied to sequencing the entire genome. The genomic DNA would be cut into pieces small enough to be handled in the laboratory, and then chemical methods would be used to determine the nucleotide sequence within each fragment. The methods are well understood, are being applied in dozens of laboratories around the world, and are a standard exercise in the education of molecular biology graduate students (33). Through February

'Polymorphisms are "common" mutations that are present in the population at a relatively high frequency. Several alternative versions of the same basic gene sequence, all equally functional, can co-exist. New mutations are defined as occurring in less than 1 percent of the population, excluding the more common polymorphic variants.

1985, the methods have been used to examine discrete and defined segments of DNA from organisms with large genomes and complete genomes from smaller organisms, sequencing a total of 4,045,305 base pairs (150)—roughly 1/800 of the human genome.

Practically, we are far away from being able to sequence the entire genome, not because the techniques are unavailable but because they are expensive and slow when applied to a task of the magnitude of sequencing 3 billion nucleotides. A meeting during the summer of 1985 considered this enormous task and estimated the cost at about $1 per nucleotide, or $3 billion overall. The effort can also be estimated in terms of technician-years, and that comes to about 200 technician-years to sequence 100 million nucleotides, which would be expected to yield about one new mutation.

Although it is conceptually satisfying, the enormous resources necessary to sequence the entire genome make it unlikely that the task will be undertaken soon. However, genomic sequencing technology is still progressing, and further automation of the process may increase its efficiency. Continued encouragement of those efforts and provision of sufficient resources to develop techniques may make it possible to sequence the entire human genome in a reasonable time. When it is done, it will probably be undertaken as part of an effort to provide more general information about human genetics, rather than to identify human mutations. If the genome is sequenced, for whatever reason, it will be a great boon to mutation studies by providing a reference for all techniques that involve sequencing or the use of restriction enzymes.

Restriction Fragment Length Polymorphisms

A method that offers a less exhaustive survey than genomic sequencing, but which is potentially more efficient, involves the use of restriction enzymes and gel electrophoresis to detect nucleotide substitutions or small deletions in the DNA (98).

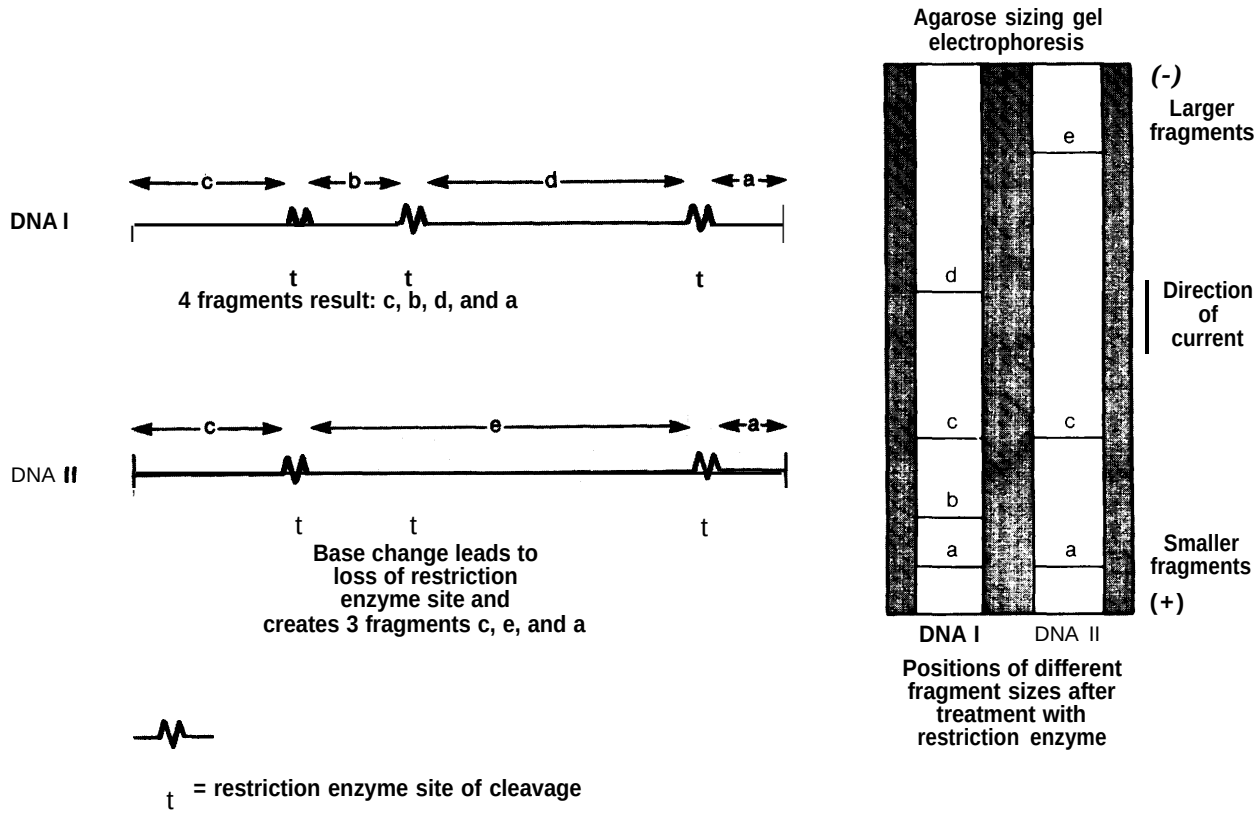
Genomic DNA is isolated from white blood cells and treated with restriction enzymes, chemicals that recognize specific sequences in the DNA and "cut" the DNA wherever such sequences occur (see fig. 7). (The number of fragments produced is determined by the frequency of occurrence of the particular enzyme recognition site in the DNA sequence.) Restriction site analysis does not, in practice, examine every nucleotide. However, the use of a set of combined restriction enzymes increases the number of restriction sites identified, allowing examination of a larger portion of the DNA, including both expressed and nonexpressed regions.

After the DNA is fragmented (see fig. 8), there are two alternative ways of proceeding, one using a gene cloning method, and the other using a noncloning, direct method.

Cloning is the process of "growing" human DNA in other organisms; it reduces the amount of DNA needed—in most cases corresponding to the amount of blood that must be drawn—but it increases the number of laboratory manipulations. Additionally, it introduces some uncertainty about the possibility that DNA changes will occur during cloning. This can be checked, but requires diligence.

In the gene cloning method, the human DNA fragments are incorporated into the genetic material of a virus called "lambda. " Lambda reproduces in the bacterium *E.coli,* and large quantities of viral DNA containing the human segments are produced. Following isolation of the cloned human DNA, each clone is treated with a set of restriction enzymes that cut it into smaller fragments. These collections of fragments are then separated by gel electrophoresis, producing distinct bands visible by gel staining methods. A visual comparison of the location of the bands derived from parental DNA and from children's DNA shows whether a mutation has occurred in a site for a restriction enzyme. The presence of a band in the child's DNA that is not present in the parents' DNA, or vice versa, indicates a nucleotide substitution or deletion restriction enzyme recognition site and suggests that a heritable mutation has occurred.

Figure 7.— Production of Restriction Fragment Length Polymorphisms Using Restriction Enzymes and Separation of Different DNA Fragment Sizes by Agarose Gel Electrophoresis



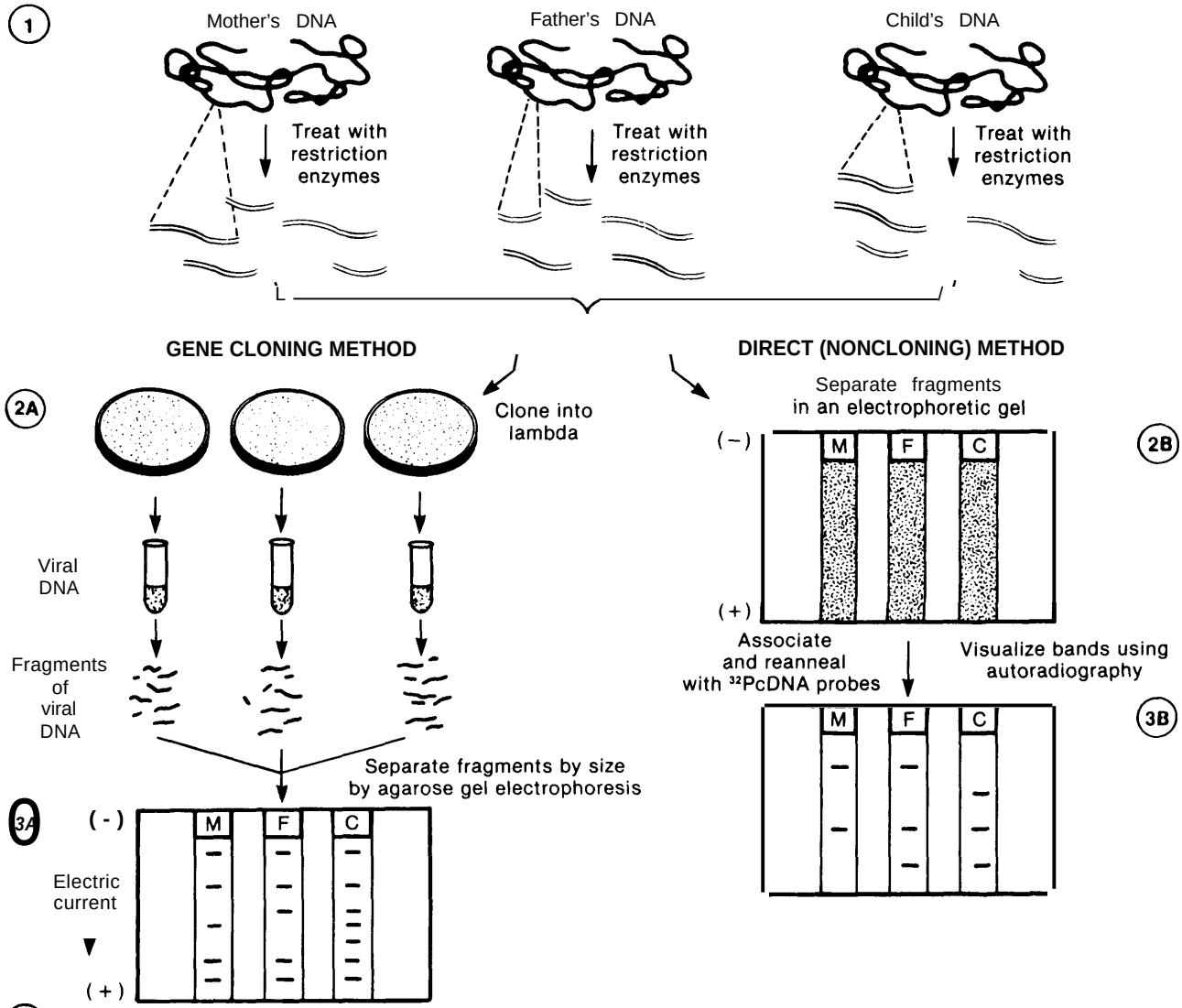
SOURCE: Off Ice of Technology Assessment

An alternative method bypasses the lambda cloning procedure, and analyzes DNA fragments directly. Fragments from the entire human genome are spread out on an electrophoretic gel, forming smears of indistinguishable bands. In order to visualize the location of particular segments of DNA and to evaluate their position relative to the equivalent band in the reference sample, the naturally double-stranded DNA is dissociated into single-stranded DNA while still in the gel and allowed to reassociate with small pieces of specific genes which are added to the gel and which are radioactive (probes). Those original DNA sequences that pair with the probes are now tagged. The DNA is then transferred to a flexible plastic membrane conserving the spatial arrangement of the DNA (a procedure known as Southern blotting) and autoradiographed; bands are detected where the tagged sequences are present. As with the cloning method above, the appearance of a new band or the disappearance of an old band suggests a nucleotide substitution or deletion in a recognition site of one of the restriction enzymes.

One= Dimensional Denaturing Gradient Gel Eiectrophoresis

A modification of the standard electrophoretic gel procedure, proposed by Fischer and Lerman (34), allows DNA to be separated not only on the basis of size, but also on the basis of sequence of nucleotides even if differences do not occur at restriction enzyme recognition sites. Double-stranded DNA dissociates into single-stranded DNA when it is heated or when it is exposed to denaturing chemicals (e.g., formamide or urea). A gradient of increasing strength of such chemicals can be produced in a gel so that DNA samples will travel in the direction of the electric cur-

Figure 8.— Restriction Fragment Length Polymorphism



1 Isolate geonomic DNA from white blood cells. Using restriction enzymes, cut DNA into double-stranded fragments of various lengtns.

Gene Cloning Method:

2A Clone each fragment into the bacterial virus lambda, infect into *E.co/i,* grown in petri dishes, and allow the virus to replicate within the bacteria. Isolate large quantities of the viral DNA, which now contains segments of human DNA, and cleave it with restriction enzymes.

3A Separate DNA fragments using agarose gel electrophoresis. Visualize bands corresponding to different sizes of fragments using fluorescent staining. Fragments found in the child's DNA samples and not in the parents' DNA may contain heritable mutations. These bands can be removed and the DNA analyzed for specific mutations.

Direct (noncloning) Method:

2B **Apply** samples of DNA fragments to an electrophoretic gel, producing smears of indistinguishable bands.

3B Dissociate DNA fragments into single strands and incubate with radioactive, single-stranded, 32P-labeled DNA probes for specific human genes. The probes hybridize with complementary sequences in the DNA and label their position in the gel. Visualize the position of bands using autoradiography. Changes in the position of bands in the child's DNA compared to the parents' DNA suggest possible new mutations. Bands can be isolated and DNA analyzed for sequence differences.

SOURCE: Office of Technology Assessment

rent, separating by size, and will also begin to dissociate as they reach their particular critical concentration of denaturing chemical. Dissociation causes the molecule to split into constituent parts, or unravel, and get stuck in the pores of the gel (84).

Every unique strand of DNA dissociates at a unique concentration of denaturant. In fact, the difference of only one nucleotide between two otherwise identical strands of double-stranded DNA of 250 nucleotides in length is enough to cause the strands to dissociate at different concentrations of denaturant chemical, and to stop traveling at different locations in the gel. Again, a comparison between the banding pattern of parents' and child's DNA analyzed in this way may identify a wide range of mutations in all regions of the DNA.

In this method, total genomic DNA is isolated from individuals and is heated to form single-stranded DNA (see fig. 9). It is then mixed with radioactive, single-stranded DNA probes that correspond to a small portion of the genome, Following incubation so that double-stranded hybrid molecules ("heteroduplexes") will form, they are treated with restriction enzymes and electrophoresed in a denaturing gradient gel. The presence of a single mismatch between a nucleotide in the "normal" probe and the corresponding segment from a person bearing a mutation makes the heteroduplex slightly more sensitive to denaturing chemicals so that such a molecule will stop traveling in the gel before the point at which a perfectly paired heteroduplex will stop. The gel is dried and exposed to X-ray film, and the *result*ing autoradiogram is examined visually for differences in banding patterns between parents' and child's DNA.

Two-Dimensional Denaturing Gradient Gel Electrophoresis

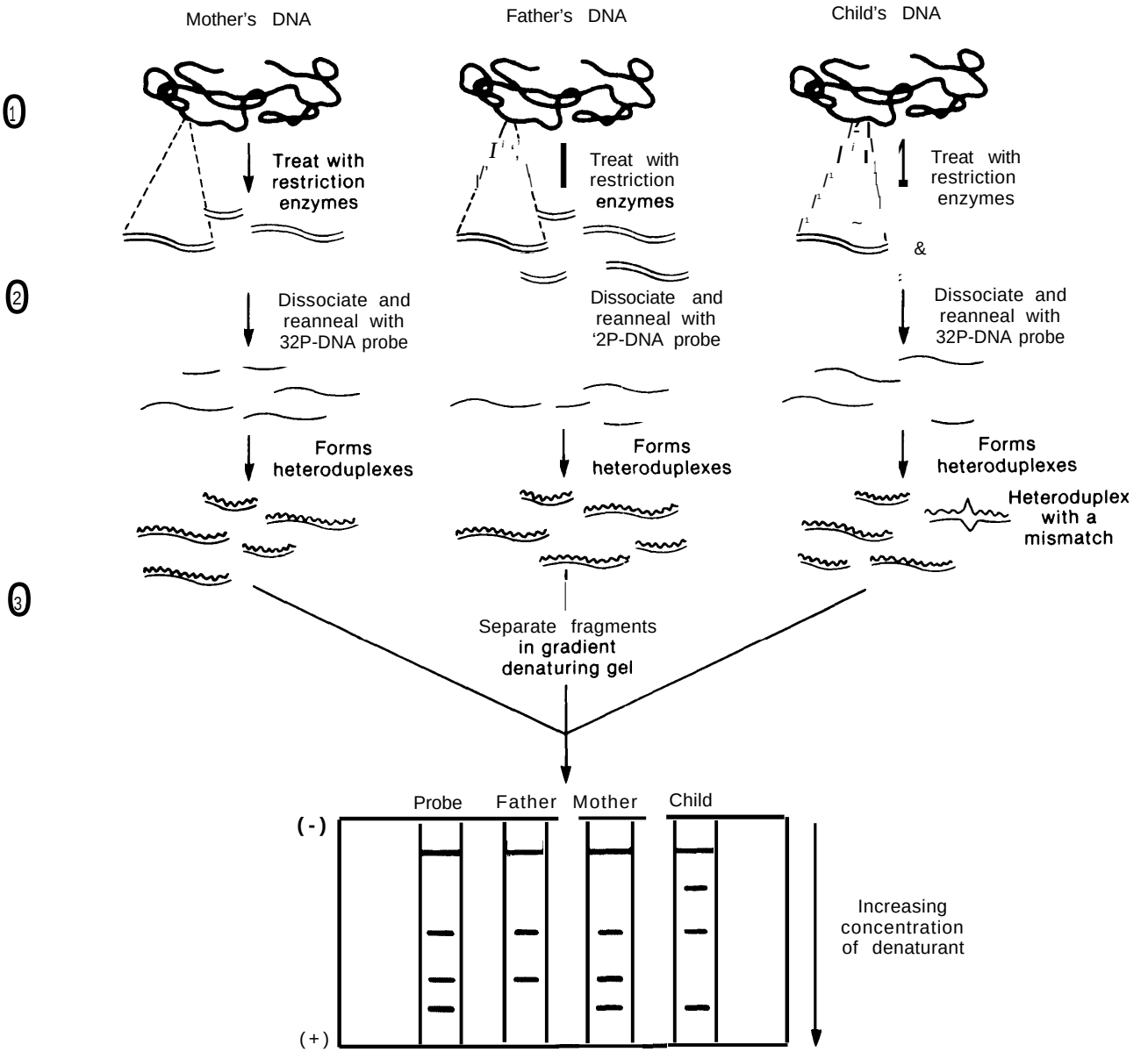
Leonard Lerman has proposed a technique whereby a two-dimensional separation of parents' and child's DNA is used to differentiate among DNA sequences common to all three members, polymorphisms in either parent which are transmitted to the child, and new mutations in the child's DNA (57). Like the two-dimensional polyacrylamide gel procedure for protein separation described above, two-dimensional denaturing gradient gel electrophoresis (2 DGGE) compares locations of spots on a gel (in this case, DNA spots) representing various types of mutations in the nucleotide sequence in expressed and nonexpressed regions of the DNA. This method separates DNA fragments on the basis of differences in base composition (or sequence), after separating them on the basis of differences in length. 2DGGE detects differences between DNA heteroduplexes by responses of their structure to gradual changes in denaturant concentration in the gel.

Genomic DNA from the parents in one sample and from the parents and children in another sample would first be treated with restriction enzymes and separated by size on separate agarose gels in a single lane (see fig. 10). The gel strips are then laid across the top of two denaturant gradient gels and the DNA is electrophoresed through the increasing gradient of denaturant. Each gel is then cut into horizontal slices, each of which contains pieces of the original homoduplex DNA of all sizes, but including only those pieces that dissociated at the same denaturing concentration ("isomelting" groups). These gel slices containing homoduplexes of parents' DNA or of parents' and child's DNA are physically removed (or collected on removable membranes) heated, and allowed to reassociate as heteroduplex molecules (see fig. **lo).**

In order to differentiate between common areas of the DNA among parents and children, another gradient denaturing gel is used to denature the mixtures of double-stranded, isomelting DNA of the parents ($M_1/M_z + F_1/F_z$) and another of the parents' DNA mixed with the child's ($M,/Mz_2 + F_1/F_2 + C_1/C_z$). Various double stranded combinations of parents' and child's DNA are produced (e.g., M_2/C_l, $F_{l,}/C_2$, etc.). The two sets of mixtures are compared in the final two gels.

Sequences that pair up perfectly represent common sequences to all three (level "i" in fig. 10). At the bottom, a narrow region, dense with heteroduplex fragments, represents the largest fraction of the sample—those that are perfectly matched between base pairs and therefore are located at the same denaturant concentration as the original homoduplexes contained in the narrow slice

Figure 9.—One-Dimensional Denaturing Gradient Gel Electrophoresis



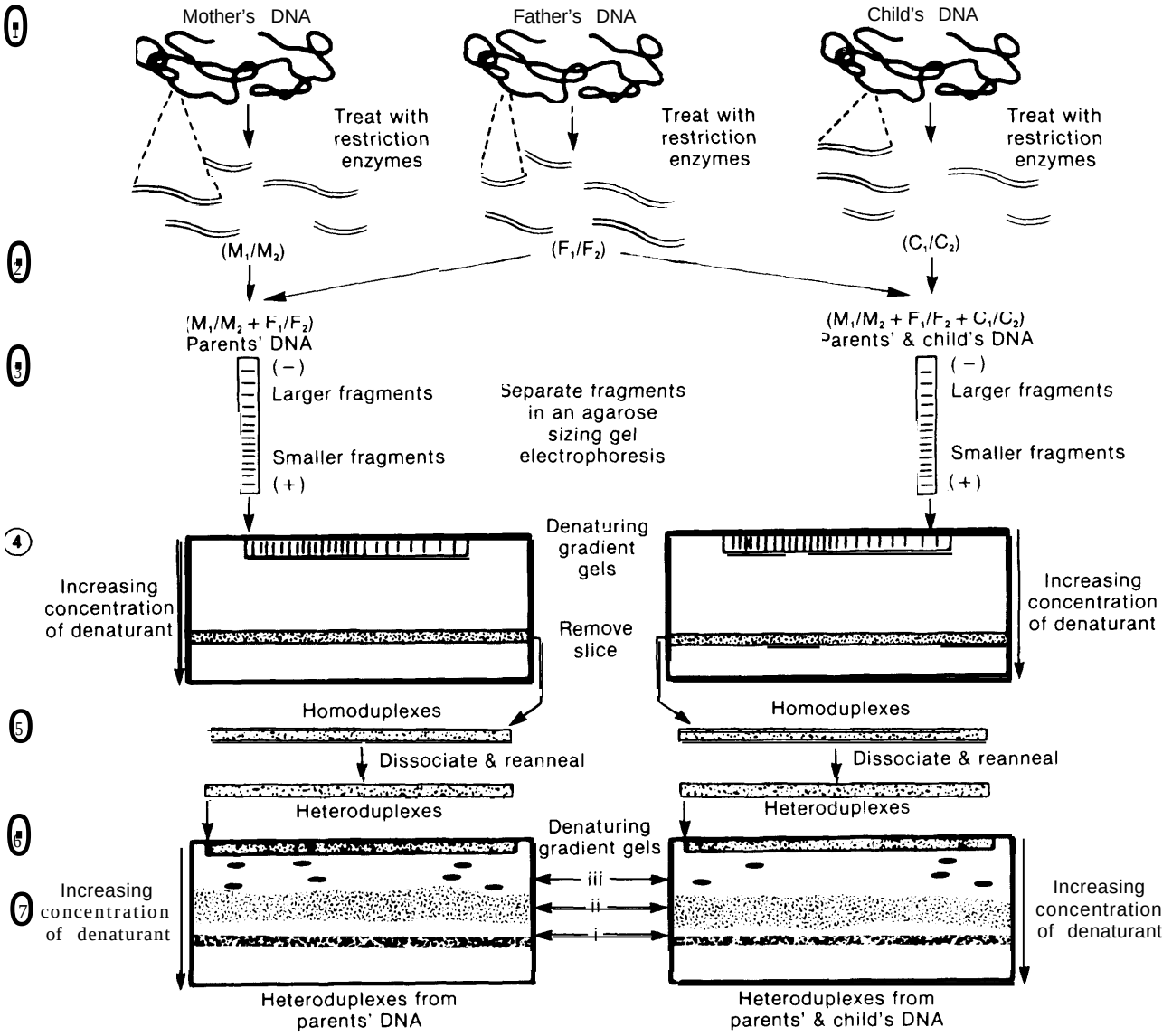
lisolate genomic DNA from white blood cells. Using restriction enzymes, cut DNA into (double-stranded) fragments of various lengths.

2 **Dissociate double-stranded DNA fragments into single strands, and reanneal in the presence of radioactive 32P-labeled, single-stranded DNA probes. Heteroduplexes form between probe and sample** DNA even if the base sequences are not perfectly complementary; if mutations are present, some of these heteroduplexes will contain mismatches.

3 Separate heteroduplex fragments in **a** denaturing gradient gel. Fragments with mismatches denature in a lower concentration of denaturant than fragments that are perfectly complementary. Visualize the position of the fragments with autoradiography. Fragments containing the child's DNA that denature sooner than the parents' fragments can be analyzed for new mutations.

SOURCE: Office of Technology Assessment.

Figure 10.—Two-Dimensional Denaturing Gradient Gel Electrophoresis



(1) Isolate genomic **DNA** from white blood cells. Using restriction enzymes, cut DNA into double-stranded fragments Of various lengths.

(2) Mix mother's and father's DNA fragments (M_1/M, + F_1/F,) in one set, and mix parents' and child's DNA fragments (M_1/M, + F_1/F, + C_1/C,) in another set.

(3) Separate fragments by length in an agarose sizing gel. Cut out a lengthwise strip of the gel, containing a range of fragment sizes, and lay the strip across the top of a denaturing gradient gel.

(4) Separate fragments along an increasing concentration of denaturant. Cut out a slice of the gel at a narrow interval of the temperature gradient, or collect a similar interval on a removable membrane.

(5) While maintaining spatial arrangement of fragments in the narrow slice, dissociate the double-stranded fragments into single strands, and reanneal them to form heteroduplexes (e. g., M_2/C_1, F_1/C_2, F_2/M_1, etc.) of various combinations of strands of the original homoduplexes (M_1/M_2, F_1/F_2, and C_1/C_2).

(6) Lay the thin strip of heteroduplex fragments arranged by size across the top of a new gradient gel, and separate in the direction of an increasing concentration of denaturants.

(7) Comparing the two final gels, the DNA spots are grouped in three distinct regions (see text for explanation). Analyze spots in the highest region of the gel for DNA fragments with possible new mutations.

SOURCE: Office of Technology Assessment.

removed from the gel or collected in the membrane.

Polymorphisms that are present in one parent or the other show up as spots that denature earlier than the perfectly matched DNAs (level "ii"). In the middle region of the gel, a diffuse distribution of fragments represents heteroduplexes with single mismatches between the two strands, causing them to dissociate earlier in the gel (at a lower denaturant concentration, occurring above the perfectly matched heteroduplexes). This region would contain inherited polymorphisms as well as new mutations, but the region would be too dense with DNA spots to distinguish individual spots.

Molecules that contain more than one polymorphism per fragment denature even earlier (level "iii") since multiple mismatches in a fragment have an additive effect on the fragments stability in the denaturing gradient. The highest area in the gel, where heteroduplexes dissociate in the lowest denaturant concentration, is where fragments with more than one mismatch would be located. These mismatches could be new mutations or inherited polymorphisms. Since the number of such fragments in the sample is likely to be small, this region would show discrete, nonoverlapping spots; DNA spots in this third level of the gel can be removed from the gel and analyzed to identify possible new heritable mutations. By comparing the two final gels, new spots should be apparent to a trained eye. These would represent single base pair changes or very small deletions and rearrangements in all regions of the DNA.

In theory, it should be possible to cut many strips of isomelting regions out of the first gel and derive an analysis of each melting region by a series of second gels. By comparing the child's DNA to his or her parents' DNA, this approach allows for large portions of the genome to be examined simultaneously.

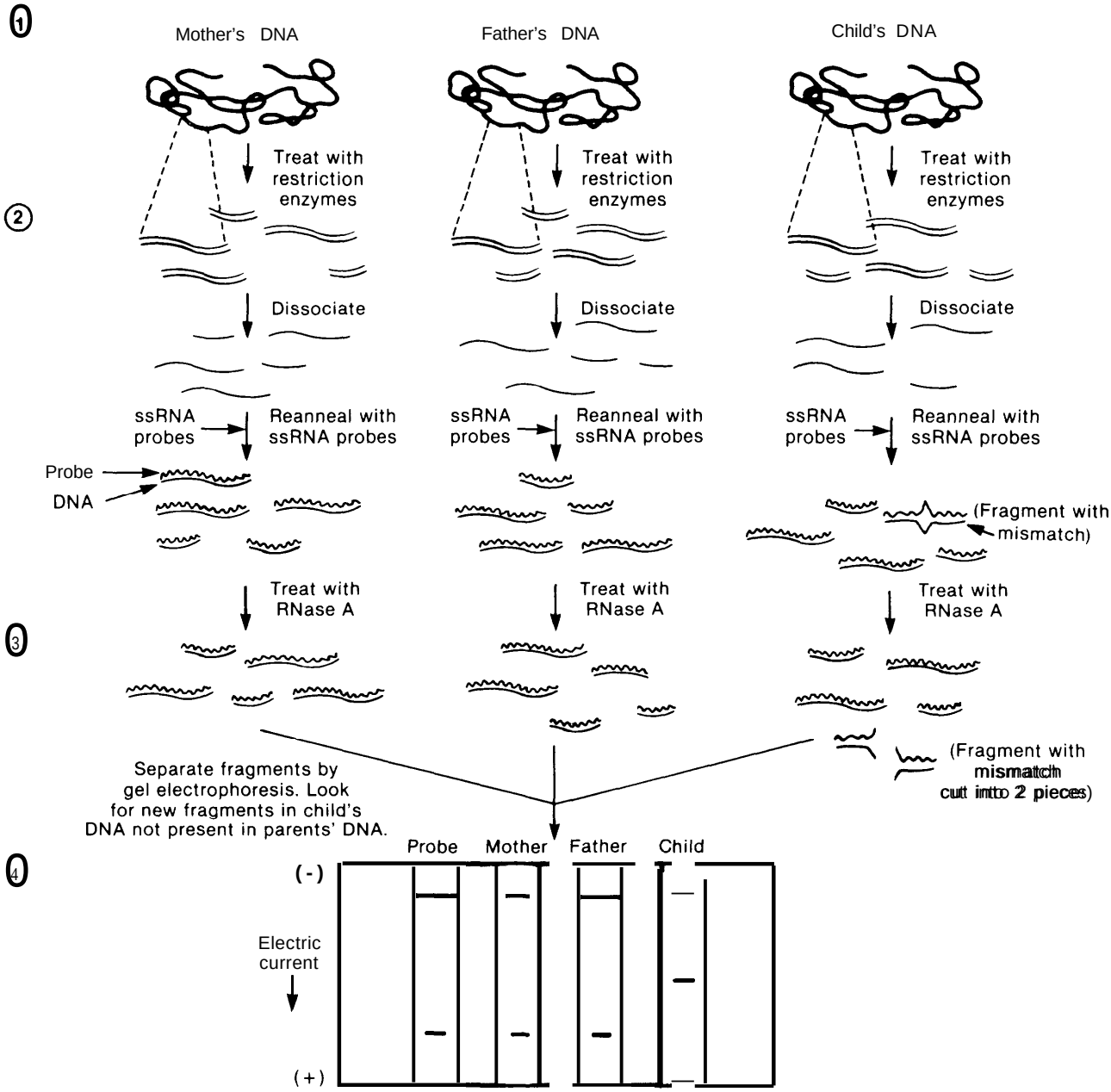
Ribonuclease Cleavage of Mismatches in RNA/DNA Heteroduplexes

A technically simpler approach to detecting mutations in cloned and genomic DNA has been proposed by Richard Myers (85,86). This technique uses an enzyme, ribonuclease A (RNaseA), that cleaves double-stranded RNA/ DNA heteroduplexes where a specific mismatch of base pairs occurs. RNaseA cleaves the RNA/ DNA molecule where a cytosine (C) in RNA occurs opposite to adenine (A) in DNA. (C normally pairs with G [guanine] and A normally pairs with T [thymidine].) The idea behind this method is similar to that of restriction enzymes, each of which cleaves DNA at specific normal sequences, except that RNaseA cleaves RNA/DNA hybrid molecules where mismatches occur. The efficiency of this approach depends on the number of different mismatches that can be recognized and cleaved. The greater the number of enzymes that can be found to cleave different mismatches, the more types of mutations can be detected. This method should detect nucleotide substitutions over a large portion of the DNA.

With this method, genomic DNA is first isolated and mixed with radioactively labeled RNA probes (see fig. 11). The mixture is heated so that the strands dissociate and then reassociate randomly with complementary strands. When RNA molecules bind with DNA molecules of homologous sequences, the result is one type of "heteroduplex." "Homologous" means that the sequence is close enough between the two molecules that they will form a stable double-stranded molecule. The presence of a single base pair mismatch does not prevent double-stranded heteroduplexes from forming, however.

The heteroduplexes are treated with RNaseA and separated according to size on a standard agarose gel. Perfectly paired molecules will be unaffected by RNaseA and will form bands on the gel that correspond to their original size. However, imperfectly paired molecules will be cut by the enzyme at the site of the C:A mismatch, and two separate fragments will result, showing up as two separate bands on the gel. For analysis, parents' and child's DNA are treated and electrophoresed separately, and the gels compared for different patterns by autoradiography to reveal the occurrence of any possible mismatches. Currently this method is applicable only for (RNA)C: A(DNA) mismatches, which represent 1 of 12 possible types of mismatches between these heteroduplexes.

Figure 11 .—Ribonuclease Cleavage of Mismatches in RNA/DNA Heteroduplexes



(1) Isolate genomic DNA from white blood cells. Using restriction enzymes, cut DNA into double-stranded fragments of various lengths.

(2) Dissociate double-stranded DNA fragments to single strands, and reanneal in the presence of single-stranded RNA (ssRNA) probes, forming RNA/DNA heteroduplexes.

(3) Treat the heteroduplexes with RNaseA, an enzyme that cuts RNA/DNA heteroduplexes where certain mismatches occur (e.g., where cytosine in RNA is mismatched with adenine in DNA).

(4) Separate heteroduplex fragments by size using gel electrophoresis. Visualize the position of fragments in the gel by autoradiography. Fragments found in the heteroduplexes containing the child's DNA, but not found in those containing parents' DNA, indicate possible new mutations.

SOURCE: Office of Technology Assessment.

Subtractive Hybridization

Detecting mutations would be much easier if it were possible to ignore the millions of nucleotide sequences that are identical in both parents and child and, instead, focus only on the relatively few sequences that are different. Currently, the best estimate for the frequency of human mutations is about 1 in 100 million base pairs, or about 30 per genome, so it is necessary to examine 3 billion base pairs from each of the parents and a child and find those 30 that differ. George Church had conceived of a method to find sequences in a child's DNA that are not present in either parent's DNA (25).

While there is no experience with Church's proposed method, and therefore, no information about its feasibility, it is a promising idea. The basis of Church's method is to view the human genome as a group of unique nucleotide sequences. In this context, "unique sequence" means a sequence that may occur only once in the entire genome. To explain what a unique sequence is, consider the shortest and longest sequence in the genome, a single nucleotide and the entire genome. A single nucleotide is not unique; A, T, G, and C occur repeatedly throughout the genome. At the other extreme, the entire genome is unique. Somewhere in between those extremes is a length of nucleotides that is long enough to be unique without being too long to handle experimentally.

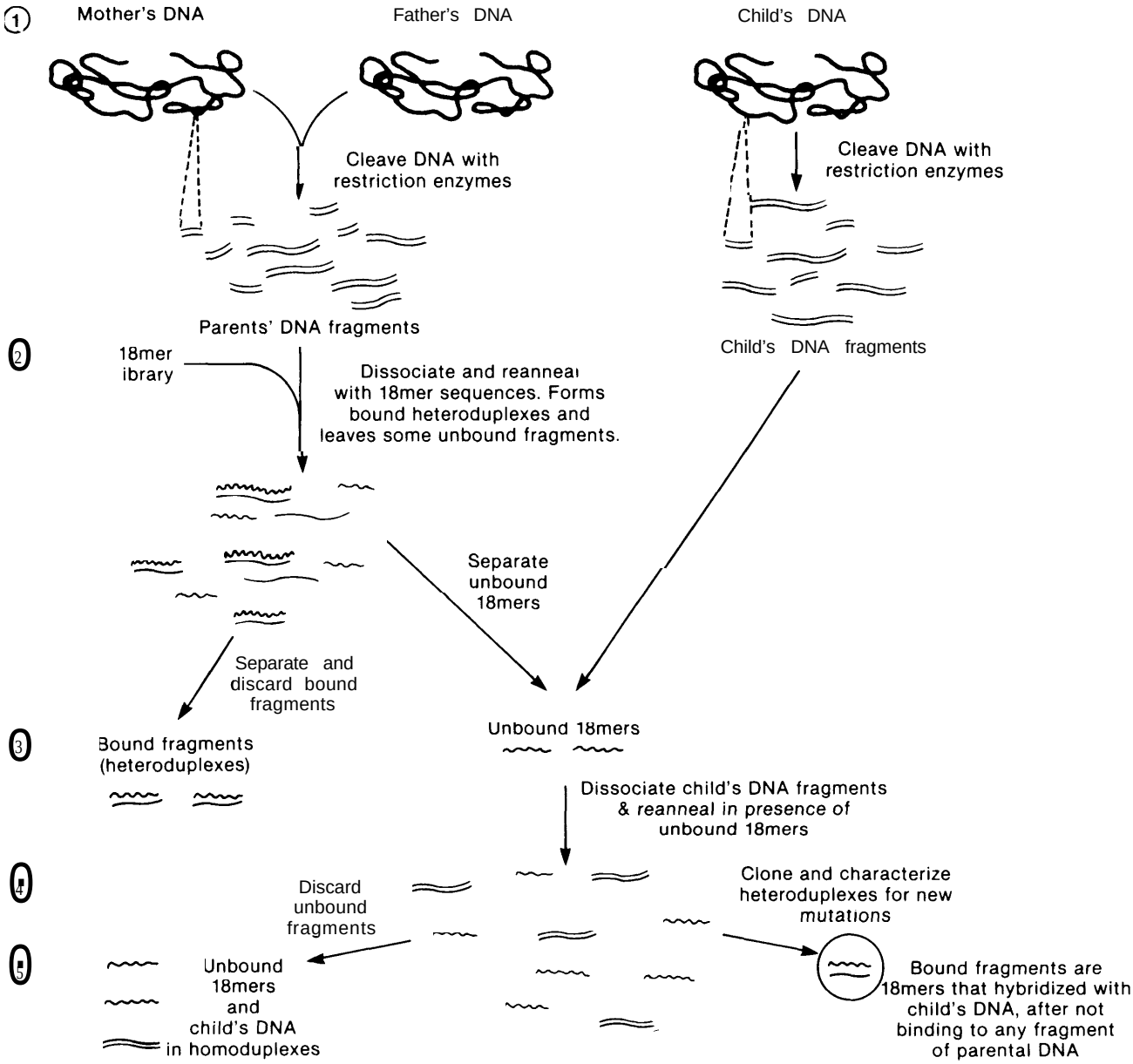
It turns out that a sequence of 18 nucleotides is long enough to be unique; on statistical grounds, any sequence of that length or greater that occurs once in the genome is not expected to occur a second time. Therefore, if every possible sequence of 18 nucleotides, or "18-mer," is synthesized, this collection of 18-mers will include sequences that together represent every possible sequence of 18 nucleotides of DNA (there are 4_{18} or 70 billion possibilities). Some proportion of these are actually present in human sequences, some are not. It is feasible to make the entire collection of 70 billion different 18-mers, and a single synthesis provides sufficient 18-mers for many experiments.

To look for mutations, DNA from parents and child is isolated and cut with restriction enzymes into relatively small lengths of 40 to 200 base pairs and dissociated into single strands (see fig. 12). This parental DNA alone is mixed with the collection of 18-mers under conditions that permit the formation of perfect hybrids: each nucleotide sequence of an 18-mer hybridizes exactly with each corresponding genomic sequence. The conditions are also such that an 18-mer that interacts with a genomic sequence complementary to just 17 of the 18 nucleotides will not form a stable hybrid. When the hybridization reaction is complete, the 18-mers that did not find a perfect match with any parental DNA sequences are left behind as single-stranded molecules, unchanged by their participation in the hybridization reaction. Because they are single stranded, they can be separated from all the other 18-mers that are hybridized to parental DNA. This unbound fraction of the mixture is retained, while the fragments of parental DNA that bind to the 18-mers are discarded.

This pool of single-stranded 18-mers which did not bind to any parental DNA sequence is then mixed with similarly prepared DNA from the child, and hybridized under the same conditions. This time, however, the molecules of interest are the sequences that bind to the child's DNA (and not previously to the parents' DNA). Any 18-mer that hybridizes perfectly with a sequence in the child's DNA will identify a sequence that is present in the child's DNA and absent from the parents' DNA. Such sequences may contain new mutations present in the child's DNA. It should then be possible to isolate and characterize this hybrid formed between the 18-mer and child's DNA, so that the mutant sequence can be determined. If mutations occur at a frequency of 1 per 100 million nucleotides, and if this technique works perfectly, it could detect 30 to 40 nucleotide sequences in the child's DNA that are not present in the parents' DNA.

This approach is the least well developed of all the ones discussed in this report, and its feasibility is unknown. If it does prove feasible, this approach could identify short sequences containing mutations in any part of the DNA, allowing further detailed study (e.g., by DNA sequencing) of the kinds of mutations that are occurring.

Figure 12.—Subtractive Hybridization



1 Isolate genomic DNA from parents' and child's white blood cells; mix parents' DNA samples together, while keeping child's DNA separate. Cleave DNA into fragments with restriction enzymes.

2 Dissociate parents' DNA fragments into single strands, and reanneal with a set of unique single-stranded 18mer sequences. 18mer sequences complementary to sequences in the parents' DNA will form double-stranded, or bound, fragments. The remainder, the unbound fraction, will remain single stranded.

3 Separate the bound fraction from the unbound fraction. The bound fraction is discarded.

4 Unbound 18mers for which no parental complement exists are allowed to hybridize with the child's DNA fragments under the same conditions of dissociation and reannealing. Any fragments of the child's DNA that hybridize with unbound fraction of 18mers represent sequences present in the child's DNA that are not present in either parents' DNA.

5 Isolate these heteroduplexes and analyze them for specific new mutations.

SOURCE Off Ice of Technology Assessment

Pulsed Field Gel Electrophoresis

If human DNA were short and simple, it could be cutup with restriction enzymes and separated electrophoretically into discrete bands, each representing a particular segment of the total DNA. However, human DNA is so long that electrophoretically separated fragments make a continuous smear of bands. If the DNA were cut in only a few places to produce only 100 or 200 bands, the pieces would be too big to pass individually through the pores of a standard electrophoretic gel. A new technique, pulsed field gel electrophoresis (PFGE) is being developed to allow separation of fragments of human DNA larger than can be separated with other electrophoretic techniques and to examine such fragments for evidence of mutations. In theory, the procedure will detect submicroscopic chromosome mutations, including rearrangements, deletions, breaks, and transpositions. At present, the method cannot handle intact human chromosomes although it works very well with smaller chromosomes from lower organisms and with small fragments of human chromosomes (157,158).

The standard process for isolating genomic DNA randomly breaks the long molecules into pieces. In the PFGE procedure, genomic DNA is treated so that random breaks are avoided. Whole cells are suspended in liquid agarose and after the agarose solidifies into a gel, enzymes are added to degrade proteins and RNA. This apparently minimizes random damage to the high molecular weight DNA. DNA is cut into exact and predictable fragments by addition of a "rare cutter, " a restriction enzyme (or group of enzymes) that cuts DNA consisting of 6 billion nucleotides, into only 3,000 pieces averaging 2 million nucleotides each (about 1/60 of an average chromosome) (126).

Ten different fragments are each labeled with a radioactive probe, so that each lane in the gel will have 10 visible bands, the most that can be analyzed per lane. The fragments are then separated on an agarose gel in which the electrical field is applied alternately ("pulsed") in perpendicular directions for 24 to 72 hours. The pulse time is adjusted, according to the particular size of the fragments, to maximize separation of the fragments. The optimal pulse time is one in which the fragments are constantly untangling and reorienting themselves. Autoradiography is the final step in the procedure. Each of the 10 fragments bound to a labeled probe appears as a visible band on the gel on exposure to X-rays; chromosomal mutations would appear as a shift in the position of the fragment containing the mutation.

Recent evidence suggests that fairly large pieces of human chromosomes can be separated, although most of the experience with the technique is limited to certain lower organisms such as yeasts and unicellular parasites. Results of experiments with DNA derived from these organisms, which have much smaller genomes than humans, suggest that chromosome mutations would be apparent if they were at least 5 percent as large as the fragment itself. The procedure would have to be modified for human chromosomes due to their larger size and complexity (127). This technique may be useful in detecting chromosome mutations that are intermediate in size between major rearrangements (observable by cytogenetic methods) and single base pair changes.

CONCLUSIONS

In a short period of time, enormous advances have been made in the ability to "read" the genetic material and to understand what it means. The techniques described in this chapter are examples of state-of-the-art molecular genetics. Some of the components of these techniques were originally developed for other purposes in genetic research, while others were designed specifically for detecting new heritable mutations. They all are successes in the sense that they propose reasonable and verifiable ways of examining human DNA for alterations in the nucleotide sequence. However, none of the DNA-based techniques is approaching the efficiency that would be needed for any of them to be appropriate for field use. As these technologies develop and information is shared among

investigators, it may be possible to design combinations of methods that complement each other and that select for those samples in which there is a greater likelihood of finding new mutations.

In general, despite impressive achievements that have been made, the technology remains too unwieldy for large-scale applications. Olson (98) likens the current situation in mutation research to the status of computer technology in the late 1950s, The spectacular improvement in computers came not from scale-up of 1950s technology but from development of the integrated circuit. He suggests that big improvements in technology will provide the appropriate methods; doing more of what we have already done is unlikely to be sufficient.

Analyzing complete sets of human DNA from large numbers of people using these methods would, at present, require an enormous commitment of resources. In considering the wisdom of devoting major resources to such research, Olson (98) emphasizes that the program's ultimate goal would have to be far more ambitious than the detection of a few bona-fide examples of new mutations. A program large enough to have a legitimate impact on public policy about environmental exposures would have to be much larger than a 5 or 10 person laboratory and be several orders of magnitude more efficient than existing restriction fragment length polymorphism technology.

One of the outstanding characteristics of current research in molecular genetics is the rapid pace of new developments and ideas; assuming this continues or accelerates, we can count on better methods to replace the proposed ones described in this report. The better ones should be able to examine at least the major portions of the genome and be practical in studying populations at higher risk for heritable mutations. They should also be able to detect a wide, if not the entire, spectrum of mutational events.

Until the next generation of techniques is developed, the best course for the next few years may be to maintain a balance between the support of current methods and the support of basic research which is the source of better ones. Maintaining a balance implies that resource-intensive efforts should not pull unwarranted amounts of resources away from developing new techniques, although no special scientific effort seems to be needed to encourage this highly innovative and rapidly paced area of research.