Massively parallel methods are not new to molecular biology: they were introduced through bacterial genetics and the screening methods that enable selection of single mutants from large numbers of cells. The introduction of DNA cloning transformed genetics: genes, formerly seen only as effects of mutations on the phenotype, are now available as DNA sequences. New techniques of DNA analysis allow mutations to be seen directly as sequence differences. These developments have transformed the genetic analysis of species, particularly of humans, in which classical genetics was greatly limited by the scarcity of good markers.

In the early to mid-1980s, the success of these molecular approaches to a wide range of genetic problems, most notably the isolation and characterization of a number of human disease genes, fuelled ambitious plans to apply sequencing and mapping on a genome-wide scale. While many groups pushed ahead, extending the scale of application of existing methods, others were motivated to develop cheaper, more readily automated methods. Existing methods are largely based on gel electrophoresis, which is a labour-intensive expensive technique and difficult to automate. The idea of using hybridization with synthetic oligonucleotides on a large scale as a way of analysing nucleic acid sequences, for 'fingerprinting', for comparing or even sequencing de novo, occurred independently to several groups.

Two strategies emerged. For fingerprinting and sequencing, the number of targets to be analysed was scaled up by spotting them robotically onto membranes, where they are hybridized with oligonucleotides applied one at a time; as 10000 clones can be hybridized at the same time on one membrane, each hybridization with, for example, a decanucleotide would generate hundreds of kb of sequence information. This has become known as format I. Despite its enormous potential, this method has been slow to take hold, possibly because hybridization with short probes is less reliable than with long ones and because the approach is suitable only for very large projects, because it requires a large number of analyses before meaningful data accumulate. The alternative is to make large arrays of oligonucleotides on a solid support, which can then be hybridized with a labelled target sequence. These devices, also known as 'DNA chips', have many potential applications. This mode of hybridization, sometimes called format II, and variations, is the subject of this review.

Large scale arrays

The problem of how to make large numbers of oligonucleotides attached to a solid support at specific sites or addresses, and attached in a way that permitted hybridization and detection, was solved remarkably quickly. Three materials were developed as solid supports: surface modified glass and polypyrrole that allow in situ oligonucleotide synthesis or post-synthetic coupling, and glass with small patches of activated polyacrylamide, to which presynthesized oligonucleotides can be attached by microinjection.

For synthesis in situ, it is necessary to direct the chemical reactions to specific addresses on the surface. This can be done with masks. For some applications, especially for the analysis of a sequence that is not known, it is desirable to have a complete set of oligonucleotide sequences of a chosen length. Such sets can be very large: there are 4^9 sequences of length 9 (e.g. 65 536 octanucleotides). 4^9 sequences can be made in only 9 coupling steps using combinatorial methods. For example, all trinucleotides can be made by a process that mimics the familiar way of writing the genetic code, in which each of the bases is coupled in turn in the rows and columns and in the same order as the triplet code is normally written. Continuing the process with stripes of ever decreasing width allows the synthesis of a complete set of any size (Fig. 1a). Channel plates can readily apply stripes at 1 mm centres and with these, an array of all octanucleotides would be 256 × 256 mm, rather larger than optimum. Some groups are testing the use of high resolution printing devices to apply the precursors or the masks. At a resolution of 300d.p.i., which is readily achieved with ink-jet technology, an octamer array would be only 20 mm square.

A different combinatorial method can be used to make an array of oligonucleotides to represent the complements of a known sequence. The number of coupling steps is equal to the number of bases in the target sequence and, thus, is the same as is needed to synthesize it in a conventional oligonucleotide synthesizer. Reagents are applied in a cell, circular or diamond shaped, and pressed against the surface of the solid support. The cell is moved along by a fraction of its diameter at each coupling step as the precursors are applied in the order in which they occur in the target sequence (Fig. 1b). The result is a set of oligonucleotides of overlapping sequence laid out in a tiling path that corresponds to their order in the target sequence. The advantage of using a circle or diamond shape is that oligonucleotides of all possible lengths are made, up to a limit which is equal to the diameter of the cell divided by the offset at each coupling step. Of course, if the cell were a square or rectangle, the result...
### TECHNICAL FOCUS

**Figure 1.** (a) Large arrays can be made by combinatorial methods. Specific chemical reactions are carried out at defined ‘addresses’ on the surface by masking areas at different stages of the synthetic cycle. The oligonucleotide precursors can be guided across the surface through channels formed by sealing an appropriate mask against the surface. Oligonucleotide chains protected by photolabile groups can be deprotected in specific regions by shining light through a mask. A protocol for synthesizing all 256 tetranucleotide sequences is shown. The first mask divides the plate into 16 columns, which are filled with precursors to A, C, G, and T in the pattern shown. After deprotection of the whole surface, the mask is turned through 90° and the precursors applied as before in the rows created by the mask. The mask is then changed for one with four channels marked by the bold lines, and the third and fourth layers are applied as shown. This process can be adapted to produce complete sets of polymers of any number of different monomers using masks with the appropriate number of channels. (b) Arrays of oligonucleotides corresponding to a full set of complements of a known sequence can be made in a single series of base couplings in which each base in the complement is added in turn. Coupling is carried out using a device that applies reagents in a defined area, which is diamond shaped in the figure. The device is displaced by a fixed movement after each coupling reaction so that consecutive couplings overlap only a portion of previous ones. The amount by which it is displaced at each step determines the length of the oligonucleotides. The diamond shape creates arrays of all oligonucleotides from mononucleotides up to the maximum length in a single series of couplings. (Reproduced with permission).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>T</td>
</tr>
</tbody>
</table>

(b) Making scanning arrays

**Figure 1.** (a) Large arrays can be made by combinatorial methods. Specific chemical reactions are carried out at defined ‘addresses’ on the surface by masking areas at different stages of the synthetic cycle. The oligonucleotide precursors can be guided across the surface through channels formed by sealing an appropriate mask against the surface. Oligonucleotide chains protected by photolabile groups can be deprotected in specific regions by shining light through a mask. A protocol for synthesizing all 256 tetranucleotide sequences is shown. The first mask divides the plate into 16 columns, which are filled with precursors to A, C, G, and T in the pattern shown. After deprotection of the whole surface, the mask is turned through 90° and the precursors applied as before in the rows created by the mask. The mask is then changed for one with four channels marked by the bold lines, and the third and fourth layers are applied as shown. This process can be adapted to produce complete sets of polymers of any number of different monomers using masks with the appropriate number of channels. (b) Arrays of oligonucleotides corresponding to a full set of complements of a known sequence can be made in a single series of base couplings in which each base in the complement is added in turn. Coupling is carried out using a device that applies reagents in a defined area, which is diamond shaped in the figure. The device is displaced by a fixed movement after each coupling reaction so that consecutive couplings overlap only a portion of previous ones. The amount by which it is displaced at each step determines the length of the oligonucleotides. The diamond shape creates arrays of all oligonucleotides from mononucleotides up to the maximum length in a single series of couplings. (Reproduced with permission).

The most sophisticated method of fabrication is an adaptation of the photolithographic methods used in the semiconductor industry, and employs novel

TIG March 1996 Vol. 12 No. 3
nucleotide precursors with a photocleavable protecting group\textsuperscript{4}. A specific pattern can be deprotected at each stage of synthesis by shining light through a mask. With spots as small as 10\textmu m, large numbers can be synthesized in a very small area. This method allows truly random access and can be used to make arrays of oligonucleotides with any sequence at any site. The maximum length of oligonucleotide that can be made depends on the stepwise yield, which for this technology is probably limited by the photodeprotection reaction. However, 20mers can be made (S. Fodor, pers. commun.) which are long enough for most applications envisaged for oligonucleotide arrays.

Reading sequences using the Watson–Crick pairing rules

All the applications of oligonucleotide arrays depend on the specificity of base pairing between the oligonucleotides and the target sequence. There have been many studies of these interactions, and some of the factors influencing duplex stability are understood, but the arrays have allowed us to study the hybridization behaviour of tens to several hundred oligonucleotides simultaneously\textsuperscript{10–12}, an enormous advantage over conventional methods. It is inconceivable that such wide ranging studies could be carried out in solution using one oligonucleotide at a time. With arrays made on a single surface, such as a standard microscope slide or a piece of polypropylene film, manipulations are very easy. Hybridization solutions can be applied under a coverslip and washings performed by dipping the slide through appropriate solutions. These are much simpler manipulations than those involved in casting, loading, running and analysing a gel. It is also relatively easy to make an apparatus that can perform all the wet processes and the data collection automatically.

Data collection

The small size of the spots of clones in format I or of oligonucleotides in format II places demands on the detection method. Radioactivity, which has advantages of high sensitivity, low background and ready availability, has been used for much of the development work both for format I and for format II, but it has the disadvantage of low resolving power. Furthermore, it can only be used after the hybridization and washing steps are complete, which rules out applications that have to be carried out in real time.

Fluorescence, which has been used in some of the studies, has major advantages: multiple colours can be used to label different sequences, resolution is high and it is possible to do real time measurements. The array can be observed as hybridization and melting reactions occur without the need to pause and expose to film or phosphor screen. The disadvantages are high intrinsic background and relatively low dynamic range. The advanced optical systems, based on CCD cameras\textsuperscript{4} or scanning confocal microscopes\textsuperscript{4}, which have been developed to overcome these problems, can image a large area quickly at high resolution and sensitivity; such methods are likely to find applications in other fields.

A novel detection method makes use of light scattering from microscopic particles that are used to tag the target sequence. Oligonucleotides are bound to a microscope slide, which is illuminated from one end, the glass acting as a wave guide; this mode of illumination generates an evanescent wave that penetrates only one half of a wavelength above the surface of the glass. Thus, only particles that are right at the surface scatter light, and particles in the solution phase do not interfere\textsuperscript{13}. This extremely simple method can also be used for real-time measurements, and could also find applications in other areas.

Detecting mutations

Analysis of point mutation is growing in importance with the discovery of increasing numbers of genes associated with inherited diseases and cancers; it is also important in following the transformation to drug resistance of infectious agents, such as HIV (Ref. 14). Hybridization with allele-specific oligonucleotides (ASOs) was one of the earliest methods for mutation screening\textsuperscript{15} and an inspiration for the wider applications, which are the subject of this commentary. In its original form, the ASOs were used as probes for restriction fragments separated by gel electrophoresis. The advantage of attaching large numbers of oligonucleotides to a solid support becomes important in cases, such as the cystic fibrosis transmembrane receptor gene, where several hundred different mutations have been reported\textsuperscript{16}.

One approach to analysing mutations is to apply the known ASOs in pairs on a single array, which is then probed with, for example, PCR amplified and labelled target\textsuperscript{17}. An alternative, which can also be used to look for mutations that have not been characterized, is to use ‘scanning’ arrays representing all positions of the target sequence (Fig. 2b). In its simple form, the analysis comprises hybridizing the target to the complements of each substring in the template. Any position which differs from that of the wild-type sequence will show up as a series of negative hybridization sites. A more complete analysis uses a set of arrays in which each base position on the target is represented four times (Fig. 2c). Each substring in the target is represented in four oligonucleotides: the wild type, and three in which the central position is replaced by one of the alternative bases\textsuperscript{18}. In this format, mutant alleles would show a negative response with the wild-type ASO but give a positive signal with their complementary oligonucleotide, which is especially important for the detection of heterozygotes.

Comparing sequences

Large genetic programmes and initiatives growing from genome analysis are producing a need for large scale sequence comparison. The need is for a simple method of comparing two nucleic acid sequences to see if they are the same or different. For example, once a region of a chromosome has been identified as one which carries a gene mapped by linkage analysis, identification of the DNA sequences associated with the gene is a long process that begins with the identification of all the coding sequences in the region. There might be tens to hundreds of candidates. Each must be isolated from large numbers of individuals to compare normal and mutant. As each coding sequence could be 1–20\textsuperscript{kb} long the sequencing task can be enormous and wasteful because most of the sequence will be identical in the mutants and in the wild type.
An advantage of arrays that distinguishes them from gels is that they are constructed on rigid materials and are reusable. Images are digitized and, therefore, the hybridization pattern of one target is easily compared with another by subtracting their two images in the computer. Hybridization due to sequence which is held in common between the two targets disappears, leaving a 'difference image' representing just that part of the sequence that differs between the two sequences. This potentially powerful method has been shown to work in model experiments with short sequences on relatively simple arrays of 256 octanucleotides (Fig. 2a).

**De novo sequence determination**

An aim of analysis by hybridization to oligonucleotide arrays is to provide a new and rapid method...
of sequence determination. The theory of the method was based on the assumption that the rules of base pairing would allow the reconstruction of the sequence of a target nucleic acid from its corresponding set of oligonucleotides, determined by the pattern of hybridization. \(^{8,10}\) The length of target that could be analysed is approximately equal to the square root of the number of oligonucleotides in the array\(^5\). Thus, an array of 65,536 octanucleotides could be used to determine sequences up to about 200 nucleotides and a million decanucleotides, a kilobase.

However, difficulties are experienced in using such short oligonucleotides in hybridization reactions: short duplexes are stable only under non-stringent conditions of high salt and low temperature, which also stabilize internal base pairing in the target. Internal pairing prevents bases from hybridization with oligonucleotides. There are a number of ways of relieving this problem. The obvious one of fragmenting the target to the size of the oligonucleotides in the array is difficult to achieve in practice. Complete arrays of longer oligonucleotides become very large, too large for existing methods of fabrication.

Enzymes and chips

In addition to applications that exploit molecular hybridization, arrays of oligonucleotides can be used in combination with enzymes. The combination can be especially useful in the field of mutation analysis, where enzymes enhance discrimination beyond that which can be achieved by hybridization alone. Polymerases require a primer and incorporate bases one at a time only if they match the complement in the template; the terminal base of the primer must also match that of the template. There are several ways in which the reaction can be used to identify the sequence or a single base at a selected site in the template strand (Refs 20, 21 and Refs therein). Ligases have similar requirements: two oligonucleotides can be joined enzymatically provided they both are complementary to the template at the position of joining\(^22\).

In solid phase minisequencing\(^22\), a tethered oligonucleotide is used to capture the target sequence at a position next to a variable base; DNA polymerase and a labelled triphosphate are added and the solution removed. The identity of the base is determined from the base incorporated. If fluorescence is used to tag the
nucleotide precursor, this method can be readily adapted to multicolour detection. Multiple capture probes can be tethered in the wells of a microtitre plate or as an array on a surface, and multiple targets can be captured on the ends of pins.

Ligases can be used in a related fashion. In this case, one oligonucleotide is tethered, the other in solution; either one or both oligonucleotides can be used to provide sequence discrimination. The use of enzymes in a solid phase format enables introduction of label through the use of tagged triphosphate precursors or oligonucleotides, thus, removing the need to label the target. This can be a great advantage when multiple targets are to be analysed. Thus, many of the methods that have been used to detect mutations using polymerases and ligases can be adapted for use with arrays. The indications are that discrimination is, indeed, greater than with hybridization alone2,24 (Fig. 3). The need for greater accuracy of mutation detection and for simple robust methods is emphasized by a recent survey of the quality of testing for common mutations in the CFTR gene in 40 European laboratories, which revealed 34 errors in a total of 360 tests25. The enzyme-based solid-state methods could provide the necessary precision, simplicity and robustness, and, as for the hybridization-based methods, the advantage of using arrays with enzymes is that they vastly increase the scale, making it possible to test several hundred positions of a target in a single analysis.

Conclusion

Oligonucleotide array technology has not yet lived up to its promise; a number of problems must be solved before it can be used routinely in genetic testing. But its promise is very high. It seems likely that DNA chips will be at the heart of a number of devices for general genetic analysis and for specialized tests. It is likely, too, that the development of these devices will engender others based on different combinatorial chemistries, with applications outside the field of genetics, for example in drug development.

References

3 Kroosko, K.R. et al. (1991) DNA Seq. 1, 175–188

E.M. SOUTHERN (EMS@BIOLOGICALS.COM) IS IN THE DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF OXFORD, SOUTH PARKS ROAD, OXFORD, UK OX1 3QU.

Meeting Reports

Noah’s ark on Fraser Island

The first HLGO comparative genome organization workshop, Fraser Island, Queensland, Australia, 3–6 December, 1995.

The workshop was attended by about 60 invited participants with research interests in the organization of vertebrate genomes. The sessions covered: updates of the status of genome maps in a range of species, vertebrate genome evolution, phenotypes, diseases and quantitative trait loci (QTL); and informatics.

The recent rapid expansion in the number and resolution of genome maps was confirmed. At first glance, the microsatellite-based maps would appear to offer little for comparative mapping. However, QTL that can be identified with such maps will make a valuable contribution to comparative genome analysis. For example, phenotypes mapped in mice or livestock will provide roles for the expressed sequence tags or genes of unknown function currently being added to the map of the human genome. Conversely, mapped human genes will be a rich source of candidates for QTL in livestock through a comparative positional candidate approach.

Johannes Weinberg (Univ. Cambridge, UK) presented data on heterogeneous chromosome painting (or ZOO-FISH) of chromosomes from several species of primates and cats with human chromosome-specific probes. These data provided an overview...