10. Although it is possible to image macromolecules under aqueous environments (*36*), the yield of stably bound complexes on the mica surface with the use of the current deposition method is insufficient for the statistical analysis described here. This low yield is probably attributable to screening, by the ionic medium, of the electrostatic interactions that hold the complexes to the mica. Although the molecules are not fully bathed in water under the humidity-controlled conditions described here, the macromolecules are likely to retain strongly bound and structurally essential water (*37*). These conditions are then less harsh than the desiccating environments required for the imaging of macromolecules by more traditional electron microscopic methods.
11. J. R. Levin, B. Krummel, M. J. Chamberlin, *J. Mol. Biol.* **196**, 85 (1987).
12. The RNA polymerase holoenzymes (Epicentre Technologies) and DNA fragments (30 nM fRB2-BstU1 DNA fragment) (*38*) were incubated for 10 min at 30°C at an ~2:1 ratio in a buffer containing 20 mM tris-Cl (pH = 7.6), 0.1 mM EDTA, 100 mM NaCl, 0.5 mM dithiothreitol (DTT), and 5% glycerol. The reaction was then diluted ~10-fold in the above buffer without NaCl or glycerol and was incubated on Parafilm at ~37°C for an additional 5 min before being deposited on freshly cleaved mica. Deposited samples were processed and imaged as previously described (*1, 3*). The C15 elongation complexes were formed after incubation of OPCs with 100 mM dinucleotide ApU and 10 mM each of adenosine triphosphate (ATP), CTP, and GTP (final concentrations) (no UTP) for 10 min at 30°C and then incubation with heparin (75 μg/ml) for 10 min before dilution and incubation on Parafilm like the OPCs. Heparin was added to trap the RNA polymerase not involved in the elongation complexes; the highly bent molecules characteristic of C15 complexes were observed in both the presence and the absence of heparin (*39*). Because of the specific sequence of the initial region transcribed from the P_L promoter, the majority (>80%) of the elongation complexes formed by this missing nucleotide procedure contained a 15-nucleotide transcript (C15 complexes) (*11, 38*).
13. D. J. Keller and C. Chih-Chung, *Surf. Sci.* **268**, 333 (1992).
14. Because all lateral dimensions are overestimated by about twice the radius of curvature of the tip in SFM images (*1, 3*), the ratios determined here are underestimates of the actual values. From the apparent width of the DNA molecules (110 ± 15 Å), the average radius of curvature of the tip is estimated to be ~75 Å (*1, 3*). With use of the lateral dimensions of RNA polymerase, obtained by electron diffraction of two-dimensional crystals as an estimate of the actual values (160 Å by 100 Å) (*15*), and use of the above estimate of the radius of curvature of the tip, the ratio of lateral dimensions should be ~(160 Å + 2 × 75 Å)/(100 Å + 2 × 75 Å) = 1.24, similar to the observed values.
15. S. A. Darst, E. W. Kubalek, R. D. Kornberg, *Nature* **340**, 730 (1989).
16. W. Tichelaar *et al.*, *Eur. J. Biochem.* **135**, 263 (1983); O. Meisenberger, H. Heumann, I. Pilz, *FEBS Lett.* **112**, 117 (1980).
17. H. Heumann *et al.*, *J. Mol. Biol.* **201**, 115 (1988).
18. The term "bending" is used here to indicate the deviation of the flanking arms of the template at both sides of the polymerase from the straight configuration, even though all the DNA is not in a double-stranded form in the region that is in contact with the polymerase.
19. H. Heumann, M. Ricchetti, W. Werel, *EMBO J.* **7**, 4379 (1988).
20. G. Kuhnke, H.-J. Fritz, R. Ehring, *ibid.* **6**, 507 (1987).
21. J. Hirsh and R. Schleif, *J. Mol. Biol.* **108**, 471 (1976).
22. R. C. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2311 (1977).
23. B. ten Heggeler-Bordier, W. Wahli, M. Adrian, A. Stasiak, J. Dubochet, *EMBO J.* **11**, 667 (1992).
24. P. Schickor, W. Metzger, W. Werel, H. Lederer, H. Heumann, *ibid.* **9**, 2215 (1990).
25. W. Metzger, P. Schickor, H. Heumann, *ibid.* **8**, 2745 (1989).
26. A. J. Carpousis and J. D. Gralla, *J. Mol. Biol.* **183**, 165 (1985).
27. A. Spassky, K. Kirkegaard, H. Buc, *Biochemistry* **24**, 2723 (1985).
28. D. C. Straney and D. M. Crothers, *Cell* **43**, 449 (1985).
29. To determine if the observed bend-angle distributions can be accounted for by random thermal fluctuations, we compared the data with a Gaussian probability distribution

$$\mathcal{P}(\theta) = (\pi\ell/P)^{1/2}\exp(-P\theta^2/\ell)$$

where $\mathcal{P}(\theta)$ is the probability of finding a given bend angle θ over a length ℓ and P is the persistence length of double-stranded DNA (*40*). Here, ℓ is assumed to be the length covered by the polymerase as determined from protection studies (ℓ = 80 bp for OPCs and 40 bp for C15 complexes) (*25, 41*). The Gaussian is centered at 0°, as would be expected for a DNA fragment possessing no intrinsic bends or curvature. The DNA within 100 bp of the transcription start site of P_L does not appear to be intrinsically bent (*42*) [also confirmed by the SFM images (Fig. 4C)]. The width of the distribution is determined by the bending rigidity of the DNA template as characterized by its persistence length, P = 210 bp at 25°C (*43*). The Kolmogorov-Smirnov test was used to compare the observed and the theoretical distributions. This test indicated that the observed data are not likely to be drawn from the theoretical distribution (significance = 0.000 for OPC data and <0.000 for C15 data).
30. U. M. Hansen and W. R. McClure, *J. Biol. Chem.* **255**, 9564 (1980).
31. B. Krummel and M. J. Chamberlin, *Biochemistry* **28**, 7829 (1989).
32. D. Rhodes and M. J. Chamberlin, *J. Biol. Chem.* **249**, 6675 (1974).
33. R. Schafer, W. Zillig, K. Zechel, *Eur. J. Biochem.* **33**, 207 (1973).
34. K. Arndt and M. J. Chamberlin, *J. Mol. Biol.* **211**, 79 (1990).
35. R. L. Novak and P. Doty, *J. Biol. Chem.* **243**, 6068 (1968).
36. C. Bustamante, D. J. Keller, G. Yang, *Curr. Opin. Struct. Biol.*, in press.
37. J. T. Edsall, in *The Proteins*, H. Neurath and K. Bailey, Eds. (Academic Press, New York, 1953), vol. 1, part B, p. 549.
38. W. A. Rees, P. H. von Hippel, A. K. Das, unpublished data.
39. W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, C. Bustamante, unpublished data.
40. M. D. Barkley and B. H. Zimm, *J. Chem. Phys.* **70**, 2991 (1979).
41. B. Krummel and M. J. Chamberlin, *J. Mol. Biol.* **225**, 239 (1992).
42. H. Giladi, M. Gottesman, A. B. Oppenheim, *ibid.* **213**, 109 (1990).
43. S. B. Smith, L. Finzi, C. Bustamante, *Science* **258**, 1122 (1992).

44. We thank S. E. Weitzel for technical assistance, G. P. Harhay for assistance with data analysis, and P. H. von Hippel, D. Erie, and A. Das for comments on the manuscript. Supported by U.S. Public Health Service (USPHS) research grants GM-32543 (C.B.) and GM-15792 and GM-29158 (P. H. von Hippel), National Science Foundation grant MCB-9118482 (C.B.), and a grant from the Lucille P. Markey Charitable Trust to the Institute of Molecular Biology. W.A.R. was a predoctoral trainee on USPHS institutional training grant GM-07759.

8 January 1993; accepted 13 April 1993

DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-Scale Sequencing

R. Drmanac, S. Drmanac, Z. Strezoska,* T. Paunesku,* I. Labat,* M. Zeremski,* J. Snoddy,† W. K. Funkhouser, B. Koop,‡ L. Hood,§ R. Crkvenjakov‖

The concept of sequencing by hybridization (SBH) makes use of an array of all possible *n*-nucleotide oligomers (*n*-mers) to identify *n*-mers present in an unknown DNA sequence. Computational approaches can then be used to assemble the complete sequence. As a validation of this concept, the sequences of three DNA fragments, 343 base pairs in length, were determined with octamer oligonucleotides. Possible applications of SBH include physical mapping (ordering) of overlapping DNA clones, sequence checking, DNA fingerprinting comparisons of normal and disease-causing genes, and the identification of DNA fragments with particular sequence motifs in complementary DNA and genomic libraries. The SBH techniques may accelerate the mapping and sequencing phases of the human genome project.

The success of the human genome project will depend on whether DNA sequencing approaches can greatly increase throughput [at least 100-fold more than the current value of ~10^4 base pairs (bp) per day per machine] and decrease cost. Strategies that may help to accomplish this task include a greatly improved method of sequencing based on the conventional automated fluorescent DNA sequencers and the advent of new sequencing technologies (*1–3*).

Any linear sequence is an assembly of overlapping, shorter subsequences. Sequencing by hybridization (SBH) (*4–10*) is based on the use of oligonucleotide hybridization to determine the set of constituent subsequences (such as 8-mers) present in a DNA fragment. Unknown DNA samples can be attached to a support

and sequentially hybridized with labeled oligonucleotides (format 1); alternatively, the DNA can be labeled and sequentially hybridized to an array of support-bound oligonucleotides (format 2). Highly discriminative hybridization is required to distinguish between perfect DNA fragment and oligonucleotide complementarity and all hybridizations exhibiting one or more nucleotide mismatches. Reliable conditions for this discrimination have recently been determined for format 1 (*5*).

The sets of *n*-mer oligomers used as hybridization probes can vary in number from several hundred to all possible combinations (65,536 for octamers), depending on the type of sequence information required. The completeness of the probe set and its design, which can vary according to such parameters as length of probe and internal (*7*) or flanking positioning (*4*) of unspecified bases, determines the kind of sequence information that can be extracted from individual DNA fragments (*7*) or libraries of fragments (*4*). Mapping information that determines clone overlap (*11*) can be obtained with 100 to 200 probes. The positioning and identification of genome structural elements (partial sequencing) (*12–14*) requires 500 to 3000 probes, and complete sequencing (*4*) requires data from 3000 or more septamer probes on three to five related genomes (*14*).

In this report, we present the results of a blind test of the SBH method on homologous DNA fragments cloned into an 8-kb M13 vector. The test consisted of sequencing by traditional methods (*15*), in one of our laboratories, three 2-kb inserts containing related variable gene segments (92 to 94% similarity) from several primate T cell receptor β loci and then resequencing by SBH, in the other laboratory, homologous 116-bp regions of these clones. The experiment was designed to determine the ability of SBH not only to produce accurate sequence information, but also to circumvent for practical reasons the SBH requirement for thousands of probes and clones in an initial test. Instead of 65,635 octamers, a reduced probe set based on gel-determined sequences was synthesized with about twice as many nonmatching as matching probes in relation to each of the 116-bp targets. The simultaneous probing of three similar sequences containing many single base mismatches within the target-probe combinations provides a stringent test of the accuracy of SBH, which justifies use of the reduced probe set.

The lists of 8-mers with more than three G + C bases and 9-mers with one to two G + C bases (*16*) occurring in the human and two rhesus monkey 116-bp DNA segments were scrambled and sent to the Argonne group along with number-coded samples of the three clones (clones 1, 7, and 8). The Argonne group was informed that within the list was a sublist that allowed unambiguous reconstruction

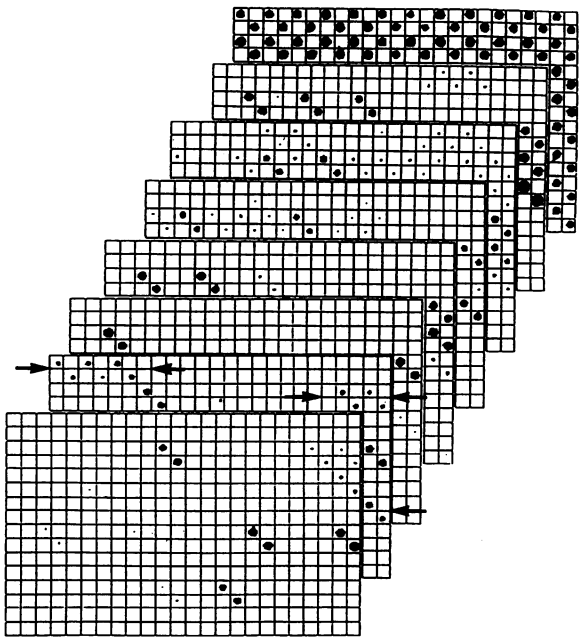

Fig. 1. Representative hybridization images obtained from one filter with different oligomer probes. The filter is made from one microtiter plate. Duplicate dots are present in the diagonally adjacent well, below and to the right. The remainder of the wells contain M13 clones with inserts of known sequence. The computer-drawn grid is included to facilitate comparison. The order of test clones (numbering from top to bottom and from left to right) is as follows: Clone 1, line 3 position 3, line 5 position 23, line 9 position 9, and line 13 position 11; clone 7, line 3 position 7, line 7 position 23, line 9 position 13, and line 13 position 13; and clone 8, line 3 position 11, line 9 position 23, line 9 position 17, and line 13 position 15. Images (from top to bottom) are as follows: Probe GCCAGCTTTC (with target in M13 vector included to calibrate the DNA amounts in the dots), probe GGGCCCAA, probe CTTAATATA, probe CGTGGCCT, probe GCTGGCCT, probe TGAGTGTC, probe TGAAGCTT, and probe ATAAAGAGT. For clarity, we selected probes with no positive controls to highlight the hybridization of clones of unknown sequence [except for probe TGAAGCTT, where six positive control clones (indicated by arrows) are visible: three in the upper left corner, and three on the right]. Other weaker signals visible in the images are at least tenfold lower than the full-match ones from the same image. The lowermost image shows the additional central positions of test clones hidden in the other images.

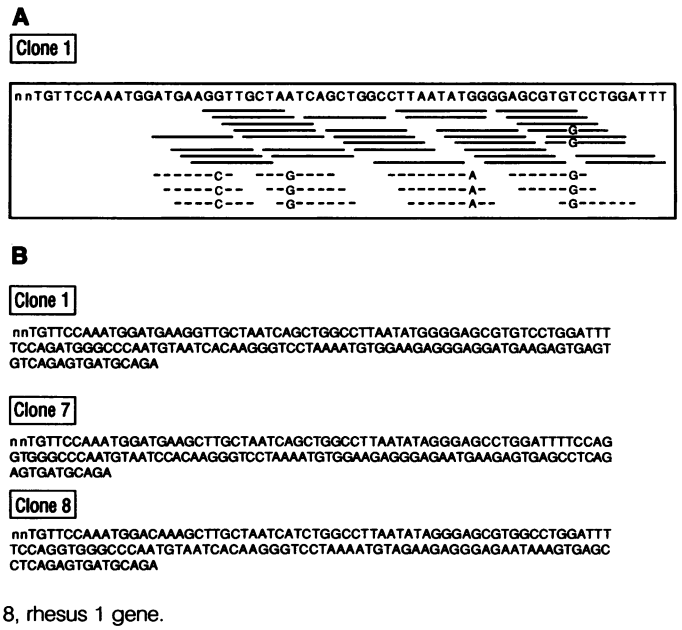

Fig. 2. (**A**) Sequence reconstruction. A part of the clone 1 sequence is depicted, showing the probes giving positive and false-positive hybridization scores (solid lines without and with G, respectively) and some of the negative scores (dashed lines). The determination of T at position 10 from right is especially instructive because it has four probes determining it as T, three as non-G, and two as G, giving odds of seven to two against G being at this position. (**B**) Sequences of three segments from T cell receptor genes determined by SBH. Clone 1, human 1a; clone 7, rhesus 2; clone 8, rhesus 1 gene.

R. Drmanac, S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, R. Crkvenjakov, Biological and Medical Research Division, Argonne National Laboratory, Argonne, IL 60439.
W. K. Funkhouser, B. Koop, L. Hood, Department of Biology, California Institute of Technology, Pasadena, CA 91125.

*Permanent address: Institute for Molecular Genetics and Genetic Engineering, Post Office Box 794, Belgrade 11000, Yugoslavia.
†Present address: Office of Health and Environmental Research, ER-72, Department of Energy, Washington, DC 20585.
‡Department of Biology, University of Victoria, Victoria, BC V8W 2Y2, Canada.
§Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195.
‖To whom correspondence should be addressed.

of a 100- to 140-bp DNA segment in at least one of the clones and that the other two sequences sent were similar. Because the list constructed from these sequences would have allowed a complete reconstruction without any of the probes being excluded, the probe information derived from a fourth similar sequence (obtained from gorilla DNA whose sample was not sent) was included in the final 272-oligomer probe list. This information made the test more challenging by increasing the number of complete sequences that could be obtained from the list.

A filter containing 96 DNA samples spotted in duplicate was prepared directly from M13 phage cultures grown in standard microtiter plates (*17, 18*). Four independent phage cultures from each of the three unknown clones were used. The remaining samples consisted of recombinant M13 clones for which the complete sequence was known (*19*). These latter DNAs served as positive and negative controls against which the hybridization signal for each probe could be calibrated. The total complexity of these control DNAs (~45 kb) was such that they provided fully matched targets for less than 50% of the supplied probes. Groups of sixteen 10- and 11-mer oligonucleotides sharing given octamers or nonamers as the informatic cores (*4*) were used as probes because shorter probes gave an insufficient signal with the available amounts of M13 DNA in the dots.

Each filter was first hybridized with a probe specific for the M13 vector so that the relative amount of DNA and hence the potential hybridization strength per dot could be measured. The filters were then successively hybridized with different probes. The discrimination values of positive signals for a given probe (average ratio of signals from matched and mismatched targets normalized in relation to variation of DNA molarity among dots) ranged from 2 to 92 for different probes, with the average value above 10 (*18*). In Fig. 1 are shown images representing the discrimination patterns of hybridization found with clones 1, 7, and 8. The successive hybridizations were stopped when 156 probes (*20*) out of the 272 supplied for the test sequence were used because sufficient data were obtained for a correct determination at each base position (Fig. 2A). On average, each base was covered by positive hybridization with more than five probes. For each clone, a continuous sequence was reconstructed (Fig. 2B), and all three test sequences were unambiguously determined.

During the hybridization of 156 probes to the three different DNA samples, 22 of 468 probe hybridization scores gave false positive results that were incompatible with the final reconstructed sequences. Half of the inconsistencies could be traced to full matches outside the targeted 116-bp sequences. Of the 11 remaining false positive results (*20*), 7 could be accounted for by signal summation from two or more single mismatch hybrids located in different regions of the clones. In only one of these seven cases, one of the mismatched hybrids was located within the 116-bp test sequence, as shown through mapping by competitive oligomer hybridization in Southern (DNA) analysis (Fig. 3A). The remaining four false positives resulted from a single mismatched target. The major cause of false positive results thus appeared to be the additive effects of two or more singly mismatched hybridizations.

In some regions, especially in a 15-bp sequence near the 5′ end of the three test sequences, we also detected an unexpected drop in hybridization intensity (Fig. 3B, left). An examination of this phenomenon (which we call "SBH compression") with model oligomer targets suggested that the discrimination of matched and mismatched hybrids within this region is retained, though at severalfold lower intensities (*21*). This hybrid instability may be caused by a secondary or tertiary DNA conformation. When longer probes specific for this target were used, a specific signal was detected (Fig. 3B, right). In some cases, nonstandard hybridization conditions or longer probes may be required to obtain unambiguous sequence data. The octamer AATATAGG fell into the false negative probe category. It did not reproducibly give the predicted hybridization pattern under standard conditions. However, when conditions were modified to detect weaker hybridizations (5-min wash at 0°C), a signal-to-noise ratio greater than five was consistently obtained.

The low percentage of false hybridizations [11/(156 × 3), or 2.3%] observed in this study is similar to that seen previously (*6*). Two considerations indicate that the observed error rate may not be a serious concern. First, software simulations suggest that as much as 10% error does not appear to affect the final sequence determination (*16*). Second, recent improvements in reconstruction algorithms have reduced the number of required heptamer probes for complete sequencing from 16,384 to about 3,000, or a fifth of all possible 7-mers (*14, 22, 23*), when the sample DNA is attached to a support. This fivefold reduction in probe repertoire allows for preselection of optimal probes in relation to hybridization errors. Thus, a well-characterized set of reagents reliable under standard hybridization conditions might be sufficient to cover all possible sequences.

The sequences of the three unknown

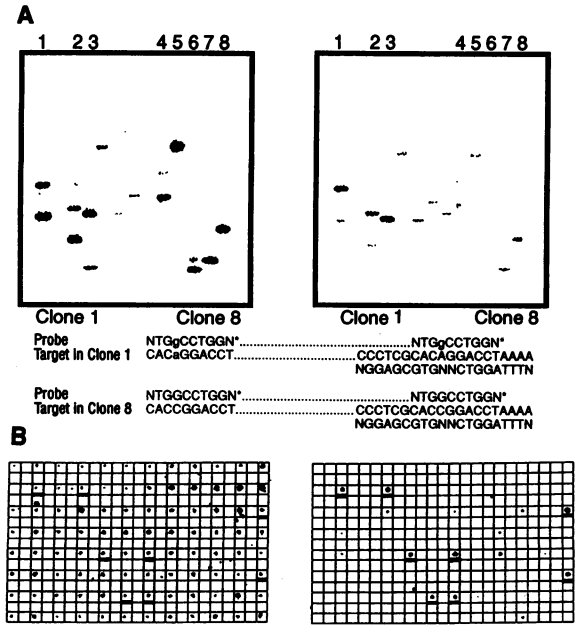

Fig. 3. (**A**) Southern blot analysis of PCR-amplified inserts of unknown clones 1, 7, and 8. Probe NTGGCCTGGN hybridized to the same filter in the absence (left) or presence (right) of the competing unlabeled oligomers NGGAGCGTGN and NCTGGATTTN specific for the fully matched or a mismatched variant of the target within sequenced segments of clones 8 and 1, respectively. Lanes 1, 2, and 3, digests of a 2-kb insert of clone 1 with restriction enzymes Pst I, Alu I, and Rsa I, respectively; lanes 4, 5, 6, 7, and 8, digests of an insert of clone 8 with Eco RI, Taq I, Alu I, Hind III, and Hpa II, respectively. The unmarked lane between lanes 1 and 2 is an Hae III digest of insert 1; no hybridization is expected because of the restriction site being within the target. The unmarked lanes between lanes 3 and 4 are various digests of insert 7; no hybridization is expected from dot blot results. The sizes of visible fragments range from 159 to 1780 bp. (**B**) Overcoming of the context-dependent depression of hybridization. (Left) Ninety-six–dot array probed with decamer NATGGATGAN; the positions of clones 1 and 7 that are expected to be positive are underlined. No higher hybridization signals in relation to expected negative control clones (all other positions) are evident. (Right) The same filter allows efficient discrimination when hybridized with 12-mer probe CAAATGGATGAA complementary to the same target but extended by two bases on the 5′ end.

DNA segments of 116 or 111 bp from human and rhesus monkey obtained by SBH were compared with the sequences obtained from gel-based traditional methods and were identical. Any systematic errors in the sequence analysis of the three DNA segments inherent to either sequencing method would have been detected.

Many of the central features of SBH technology can also be used in conjunction with standard gel-based sequencing methods. Subclones that are first ordered and partially sequenced by SBH can then be completely sequenced by traditional methods. Knowing the order and orientation of the subclones before sequencing can markedly reduce the overall redundancy commonly required in strategies based on random sequencing. This strategy also makes it possible to eliminate the added redundancy from the sequencing of overlapping cosmid clones. In the error-checking phase of the sequencing process, two different sequencing methods provide a much better estimate of real error and thus provide greater confidence in the final sequence.

By combining SBH and one-pass gel sequencing methods, overall throughput and accuracy can be improved by at least an order of magnitude. SBH can be used to fill in gaps and check errors. It is possible that SBH could itself become a primary method of genomic sequencing except for segments consisting of multiple tandem repeats. In addition, because of the combinatorial and parallel features of SBH, several similar genomes could be readily sequenced simultaneously (*14*). Furthermore, SBH is an ideal method for identifying genes, repeats, and motifs in chromosomes and cDNAs as well as for study of the differences among related populations and species (*12, 14*). The usefulness of partial sequences produced by gel methods for identifying expressed genes has already been shown (*24*).

Filter hybridization offers the ability to simultaneously analyze thousands of samples. We have shown that there is no biochemical or instrumental limitation in extending the number of DNA dots from 192 to 13,496 on a single 8- by 12-cm filter (*17*), and hybridizations to 100 such filters require only a little more effort than for a single filter. In addition, an array format allows miniaturization to a "sequencing chip" by spacing of the samples (oligonucleotides or genomic fragments) on a scale of micrometers (*5, 9, 25*). It is expected that SBH will contribute to the development of rapid and inexpensive techniques for mapping clones, facilitate DNA sequence analysis, and extend the applicability of DNA fingerprinting for diagnostics.

REFERENCES AND NOTES

1. T. Hunkapiller, R. J. Kaiser, B. F. Koop, L. Hood, *Science* **254**, 59 (1991).
2. J. H. Jett *et al.*, *J. Biomol. Struct. Dyn.* **7**, 301 (1989).
3. S. M. Lindsay and M. Philipp, *Genet. Anal. Tech. Appl.* **8**, 8 (1991).
4. R. Drmanac and R. Crkvenjakov, Yugoslav patent application 570/87 (1987); R. Drmanac, I. Labat, I. Brukner, R. Crkvenjakov, *Genomics* **4**, 114 (1989).
5. R. Drmanac, Z. Strezoska, I. Labat, S. Drmanac, R. Crkvenjakov, *DNA Cell Biol.* **9**, 527 (1990).
6. Z. Strezoska *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10089 (1991).
7. W. Bains and G. C. Smith, *J. Theor. Biol.* **135**, 303 (1988).
8. Y. P. Lysov *et al.*, *Dokl. Akad. Nauk SSSR* **303**, 1508 (1988); K. R. Khrapko *et al.*, *FEBS Lett.* **256**, 118 (1989); P. A. Pevzner *et al.*, *J. Biomol. Struct. Dyn.* **9**, 399 (1991).
9. K. R. Khrapko *et al.*, *J. DNA Sequencing Mapp.* 1, 375 (1991).
10. E. M. Southern, United Kingdom Patent Application GB 8810400 (1988); E. M. Southern *et al.*, *Genomics* **13**, 1008 (1992).
11. A. Poustka *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 131 (1986).
12. R. Drmanac *et al.*, *Electrophoreses, Supercomputers and the Human Genome* (World Scientific, Singapore, 1991), pp. 60–74.
13. H. Lehrach *et al.*, in *Genome Analysis, Genetic and Physical Mapping*, K. E. Davies and S. M. Tilghman, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1991), vol. 1, pp. 39–81; G. G. Lennon and H. Lehrach, *Trends Genet.* **7**, 314 (1991); S. Meier-Ewert *et al.*, *Nature* **361**, 375 (1993).
14. R. Drmanac and R. Crkvenjakov, *Int. J. Genome Res.* 1, 59 (1992).
15. R. K. Wilson *et al.*, *Genomics* **13**, 1198 (1992); B. F. Koop *et al.*, *BioTechniques*, in press.
16. R. Drmanac, I. Labat, R. Crkvenjakov, *J. Biomol. Struct. Dyn.* **8**, 1085 (1991).
17. R. Drmanac *et al.*, *Electrophoresis* **13**, 566 (1992).
18. Clones were grown in *Escherichia coli* JM101 in 2× YT medium (10 g of yeast extract, 16 g of Bacto-tryptone, and NaCl (5 g per liter). Plates (96-well) were shaken at 350 rpm at 37°C, and bacteria were settled by centrifugation at 3800 rpm or by standing for 3 days. We spotted supernatants by touching with flat-tipped stainless steel pins and transferring with printing on GeneScreen (New England Nuclear) on three sheets of 3 MM filter paper soaked in 0.5 M NaOH. Five to ten offprints of samples from each well were made into the same dots for transfer of a sufficient amount of M13 DNA. Further filter treatment, probe labeling, and hybridization procedures were done as described (*6*). Washing was for various times (from 5 min to overnight) and at 0 to 16°C, depending on the expected hybrid stability; the largest group of probes required 3 hours of washing at 16°C. Removal of probes was at 37°C overnight in the hybridization buffer. The filters were exposed against Storage phosphor plates for 2 to 18 hours and were scanned in a Phosphor-Imager (Molecular Dynamics, Sunnyvale, CA). The data were quantified by Image Quant and with an adaptation of Excel spreadsheet software, providing relative intensities and discrimination values for each dot-probe pair corrected for variation in the amounts of DNA (*6*).
19. In addition to the test M13 clones, 56 M13 clones with inserts of 1 to 2 kb of known sequence (obtained from R. Gibbs and T. Caskey, Baylor College of Medicine, Houston, TX, and E. Chen, Genentech, San Francisco, CA) were used.
20. The list of 156 probes used to reconstruct sequences and their hybridization scores obtained with the test clones is as follows [positive scores (see text) with clones 1, 7, and 8 are indicated by the respective numeral; the absence of a numeral indicates a negative score]: AAATGTGG 1, 7; AAAGAGTG 8; AAAGCTTG 8; AATATAGG 7, 8; AATATGGG 1; AATCAGCT 1, 7; AAGAGTGA 1, 7, 8; AAGTGTCC; AAGCTTGC 7, 8; AAGGTTGC 1; AAGGGTCC 1, 7, 8; ATATAGGG 7, 8; ATATGGGG 1; ATAGGGAG 7, 8; ATCATCTG 8; ATCTGGCC 8; ATGAAGAG 1, 7; ATGAAGCT 7; ATGAAGGT 1; ATGTCCTG; ATGGGCCC 1; ATGGGGAG 1; ACAAAGCT 8; ACAAGGGT 1, 7, 8; ACTGTTCC; ACGAAGAG; AGAAGAGG 8; AGAGTGAG 1, 7, 8; AGAGGGAG 1, 7, 8; AGAGGGGG; AGTCAGCT; AGTGAGCC 7, 8; AGTGTCAG 1; AGCTTGCT 7, 8; AGCCTGGA 7; AGCGTGTC 1; AGCGTGGC 8; AGGATGAA 1; AGGTTGCT 1; AGGGAGAA 7, 8; AGGGAGCA; AGGGAGCC 7; AGGGAGCG 8; AGGGAGGA 1; AGGGTCCT 1, 7, 8; AGGGGGGA; TAATATGG 1; TAATCAGC 1, 7; TATAGGGA 7, 8; TATGGGGA 1; TAGAAGAG 8; TAGTCAGC; TTTTCCAG 1, 7, 8; TTTCCAGA 1; TTTCCAGG 7, 8; TTTGCAGG; TTCCAGAT 1; TTCCAGGT 7, 8; TCATCTGG 8; TCACAAGT; TCACAAGG 1, 7, 8; TCAGCTGG 1, 7; TCTGGCCT 8; TCCAGATG 1; TGAAGAGT 1, 7; TGAAGCTT 7; TGAGTGTC 1; TGAGCCTC 7, 8; TGTAGAAG 8; TGTCAGAG 1; TGTCCTAA; TGTCCTGG 1; TGTGGAAG 1, 7; TGCTAATC 1, 7, 8; TGCTAGTC; TGGAAGAG 1, 7; TGGCCTTA 1, 7, 8; TGGCCTGG 1, 8; TGGGGAGC 1; CAAAGCTT 8; CATCTGGC 8; CAGCTGGC 1, 7; CAGGTGGG 7, 8; CTAGTCAG; CTTGCTAA 7, 8; CTGGATTT 1, 7, 8; CTGGCCTT 1, 7, 8; CCAGATGG 1; CCAGGTGG 7, 8; CCTGGATT 1, 7, 8; CCCAATGT 1, 7, 8; CGAAGAGT 7, 8; CGTGGCCT 1, 8; GAATGAAG 1, 7; GAAGAGTG 1, 7; GAAGAGGG 1, 7, 8; GAAGCTTG 7; GAAGGTTG 1; GATTTTCC 1, 7, 8; GATTTTGC; GACGAAGA; GAGTGAGT 1; GAGCATGT; GAGCCTCA 7, 8; GAGCCTGG 7; GAGCGTGT 1; GAGCGTGG 8; GAGGATGA 1; GTAGAAGA 8; GTTGCTAA 1; GTCCTAAA 1, 7, 8; GTGTCAGA 1; GTGTCCTG 1; GTGGGCCC 7, 8; GCAGGTGG; GCTAATCA 1, 7, 8; GCTAGTCA; GCTGGCCT 1, 7; GCCTGGAT 7, 8; GCGTGTCC 1; GCGTGGCC 8; GGATTTTC 1, 7, 8; GGATTTTG; GGAGCATG 1, 7, 8; GGAGCCTG 7; GGAGCGTG 1, 8; GGAGGATG 1; GGTTGCTA 1; GGTCCTAA 1, 7, 8; GGCCTTAA 1, 7, 8; GGCCTGGA 8; GGGAGAAT 7, 8; GGGAGCAT; GGGAGCCT 7; GGGAGCGT 1, 8; GGGCCCAA 1, 7, 8; GGGGAGCG 1; AAATGTAGA 8; AATCATCTG 7, 8; ATAAAGAGT 8; ATTTTCCAG 1, 7, 8; ATTTTGCAG; ATGTAGAAG 8; TAAAATGTA 8; TAAAGAGTG 8; TAATATAGG 7, 8; TAATCACAA 1, 7, 8; TTAATATAG 7, 8; TTAATATGG 1; TTGCTAATC 1, 7, 8; TGTAATCAC 1, 7, 8; TGGATTTTC 1, 7, 8; TGGATTTTG; CTAAAATGT 1, 7, 8; CTAATCATC 8; CTTAATATA 7, 8; CCTTAATAT 1, 7, 8; GAGAATAAA 8. Eleven false positive scores are included as follows: With clones 1, 7, and 8, probes ATGTCCTG and GGAGCATG; with clones 7 and 8, probe CGAAGAGT; and with clone 1 probes TGGCCTGG, CGTGGCCT, and GAATGAAG. An additional nine unscorable probes were tested but are not listed (reasons for noninclusion: the full-match targets in M13 or in surrounding inserts obscuring the score; the strong hybridization of all clones, control and unknown, indicating vector mismatch origin; and single test clone false-positives with no full-match controls available, that is, specific mismatch hybridization above background).
21. T. Paunesku and R. Crkvenjakov, unpublished results.
22. R. Drmanac, in *Genome Mapping and Sequencing* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), p. 318.
23. I. Labat and R. Drmanac, unpublished results.
24. M. D. Adams *et al.*, *Science* **252**, 1651 (1991); A. S. Wilcox *et al.*, *Nucleic Acids Res.* **13**, 1837 (1991).
25. R. Drmanac and R. Crkvenjakov, *Scientia Yugosl.* **16**, 97 (1990); S. P. A. Fodor *et al.*, *Science* **251**, 767 (1991); M. Eggers *et al.*, in (*22*), p. 111.

26. We thank M. Stodolsky of the Department of Energy, Office of Health and Environmental Research, for monitoring the blind test and aiding in its design. Supported by U.S. Department of Energy, Office of Health and Environmental Research, under contract W-31-109-ENG-38.

7 December 1992; accepted 15 March 1993