Double-stranded DNA arrays: next steps in the surface campaign

Robert Carlson and Roger Brent

It's only eight years since surface photoaddressable synthesis of oligonucleotides and oligopeptides was first devised[1]. But in that time, the technology has advanced with astonishing speed, addressing an ever-increasing array of applications, from expression analysis to genotyping and mutation screening. In this issue, Church and colleagues[2] take DNA chips to another level, by synthesizing double-stranded DNA arrays accessible to proteins. Their system could allow the systematic study of DNA–protein interactions, transforming the analysis of eukaryotic transcriptional control.

Current synthetic oligonucleotide arrays typically consist of single-stranded 20- to 30-mers affixed to glass surfaces by oligoethylene glycol linkers. These arrays have many uses, such as monitoring the presence and abundance of mRNAs in populations[3], assessing sequence variation in HIV-1[4], determining the relative fitness of yeast strains haploid at a given locus in a population of diploids[5], and identifying the complete set of crossovers in individual meioses in yeast[6].

In their new approach, Church and colleagues display longer single-stranded oligonucleotides (up to 40-mers), immobilized onto the surface via their 3′ ends. To these, they hybridize 16-mer priming oligonucleotides, extending these primers with Klenow fragments, thereby converting the immobilized DNA into a double-stranded form. By incorporating a fluoroescent label into the DNA as a marker, they show that this double-stranded DNA is accessible to proteins—in this case, a restriction enzyme and a methylase.

In the first experiment, an *Rsa*I cleavage site was incorporated into the DNA strand between the label and the glass. Treating the DNA with *Rsa*I eliminated the fluorescence, showing that the enzyme bound and cleaved at the correct site. In the second experiment, when a GATC site was similarly incorporated, treating the DNA with *dam* methylase and *Dpn*I released the fluorescence, demonstrating that the methylase modified, and the *Dpn*I cleaved, the GATC site.

The authors note that the ability to make double-stranded DNA arrays should enable the high-throughput study of DNA–protein interactions. Because of the marvelous way genomic techniques tend to simplify biological information—turn complex quantitative changes into qualitative ones—such studies could lead to substantial new insights into how eukaryotes regulate gene expression. In the near term, arrays that contain all possible permutations of a site at each position should help in determining for known proteins the preferred nucleotides at each position within the site. Particularly when repeated with mutant proteins that bear lesions in possible contact amino acids, such studies could rapidly pinpoint individual contacts between nucleotides and amino acids. Because of the potential speed of such studies, even small differences in the binding specificity among proteins in the same family may be detected—differences that, for most newly discovered eukaryotic DNA-binding proteins, have been poorly characterized. In the longer term, larger arrays that contain a fair representation of possible protein binding sites might help in defining DNA-binding specificities for uncharacterized proteins.

Before this technology allows such large-scale experimentation, two technical issues may need to be addressed. First, synthesis of single-stranded oligonucleotides on solid surfaces is inefficient, with per-nucleotide synthesis efficiencies thought to be only 92–96%[7,8]. Thus, for a 40-mer, only 4–20% of the sequences on a chip will be of the desired length and sequence. It may take thoughtful planning to devise protein binding experiments that can produce significant answers in the presence of so many competing related sequences. Second, there is no reason to think that all of the single-stranded oligos are accessible to the primers and polymerases needed to make the second strand. The experiments described in this paper all required enzymatic reactions on double-stranded DNA to produce output; DNA that stayed single stranded would not show up.

Economic issues may also slow development of these methods. The addressable synthesis of oligonucleotides is currently an expensive and proprietary technology. Its cost and relative inaccessibility may spur development of alternative methods to manufacture double-stranded DNA arrays, just as it did for DNA arrays for gene expression monitoring[9,10]. Here, for example, an alternative route to the construction of homogenous double-stranded DNA arrays might be based on the techniques for fabrication by hybridization and ligation described by Ben-Yoseph and his co-workers[11] (see Fig. 1B). These investigators attached single-stranded oligonucleotides to gold supports by a thiol linkage, and then to those oligos, hybridized and ligated double-stranded DNAs with the appropriate complementary ends.

Another possible high-throughput approach would be to move beyond the display of DNA in two-dimensional arrays. Existing technology allows fabrication by hybridization to "zero-dimensional" structures—beads bearing different oligonucleotides[12]. Beads with

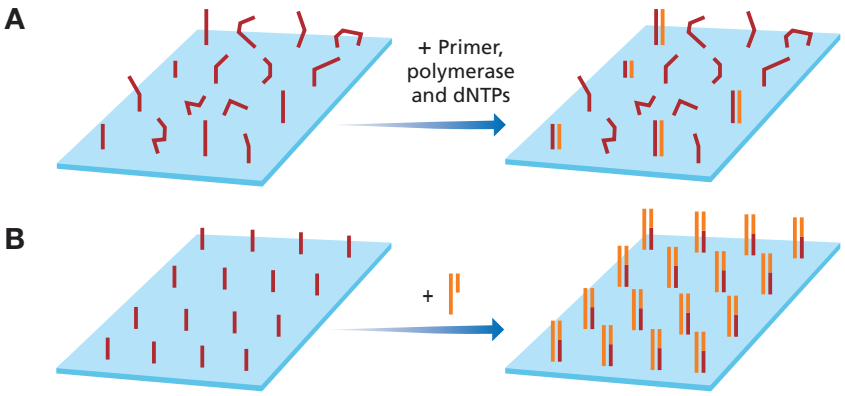

Figure 1. Alternative approaches for making double-stranded DNA arrays. (A) Upper panel: Single-stranded oligonucleotides synthesized on a glass surface[1]. Not all the strands are of unit length or accessible to the primer and polymerase, thus, some strands may not be converted into double-stranded DNA of the the desired length and sequence. (B) 12-mer oligonucletides coupled to a solid support. Hybridization to double-stranded DNA with a complementary single-stranded end, followed by treatment with DNA ligase, results in surface-coupled double-stranded DNA[10].

Robert Carlson and Roger Brent are at The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704 (brent@molsci.org).

diameters of 12 µm have larger surface areas than 20 × 20 µm patches of flat surface, even if their accessible pores are excluded. It should be now be possible to manipulate such beads in solution, and to screen them for binding-dependent fluorescence by flow cytometry. In the near future, it should become possible to manipulate beads in microfabricated fluidic devices that can sort and perform subsequent biochemical analysis on specific beads. Several million beads, arrayed serially within a serpentine channel in a microfabricated device, would occupy only as much space as the roughly 10,000 different oligos displayed on current two-dimensional arrays.

A more radical approach would be to use DNA itself as a structure that supports individual double-stranded sequences for binding assays. Nadrian Seeman and his co-workers have been developing means to self-assemble DNA into "polyhedral" three-dimensional structures (e.g., see ref. 13). Recently, they reported self-fabrication of significantly more complex structures, whose main elements each consisted of two antiparallel double helices rigidly connected by two Holliday-type junctions. They assembled these structural elements into large (micron-length), heterogeneous (multisubunit), two-dimensional crystals by hybridization of the elements' 6-mer sticky ends[14]. Significantly, in some experiments, these researchers used derivatives of these structural elements that displayed double-stranded DNA hairpins. There seems no reason such double-stranded DNA side chains should not contain binding sites. Given design rules for such crystal subunits, investigators could fabricate arrays in their own labs.

How soon these more exotic approaches may become possible is hard to guess. It's probably a good bet that double-stranded DNA arrays on surfaces will be the most accessible technology in the near term, and that these surface arrays will be good enough to enable high-throughput studies of DNA–protein interactions. In the longer term, the ability to draw on more than 40 years of nucleic acid enzymology will create an unparalleled ability to program the assembly of DNAs into structures. It's likely that a great deal of interesting biology will emerge from such assemblies. In particular, regularly spaced double-stranded DNA displayed by self-fabricated DNA structures could provide a regular array of protein binding sites. These binding sites can be used to form regular arrays of chimeric proteins that contain the appropriate DNA-binding domains. Because most proteins can be expressed as such chimeras, the ability to array them in quasi-crystals with regular spacing and orientation could facilitate their high-throughput structural analysis. Beyond these biological applications, we note that Braun et al.[11] have addressed double-stranded DNA to single-stranded oligonucleotide connected to gold electrodes on glass, then deposited silver on the double-stranded DNA to make wires[11]. Such fabrication ability could lead to hybrid DNA–electrical devices such as NMR coils.

For now, Church and his co-workers have demonstrated the promise of DNA–surface technology to study DNA–protein interactions. But the eventual impact of these studies on other aspects of biology and engineering may be even more important.

1. Fodor, S. et al. *Science* **251**, 767–773 (1991).
2. Lockhart, D.J. et al. *Nat. Biotechnol.* **17**, 573–577 (1999).
3. Bulyk, M.L. et al. *Nat. Biotechnol.* **14**, 1675–1680 (1996).
4. Kozal, M.J. et al. *Nat. Med.* **2**, 753–759 (1996).
5. Giaever, G. et al. *Nat. Genet.* **21**, 278–283 (1999).
6. Winzeler, E.A et al. *Science* **281**, 1194–1197 (1998).
7. McCall, G.H. et al. *J. Am. Chem. Soc.* **119**, 5081–5090 (1997).
8. Southern, E.M. et al. *Nat. Genet. (Suppl.)* **21**, 5–9 (1999).
9. Maskos, U. & Southern, E.M. *Nucleic Acids Res.* **21**, 2267–2268 (1993).
10. DeRisi, J.L. et al. *Science* **278**, 680–686 (1997).
11. Braun, E. et al. *Nature* **391**, 775–778 (1998).
12. Brenner, S. & Lerner, R.A. *Proc. Natl. Acad. Sci. USA* **89**, 5381-5383 (1992).
13. Chen, J. & Seeman, N.C. *Nature* **350**, 631–633 (1991).
14. Winfree, E. et al. *Nature* **394**, 539–544 (1998).