DOI:
10.1038/nrg2258

MILESTONE 7

Deciphering the code

A fundamental cornerstone of the era of "genetic engineering" was the development of technology that allowed researchers to determine a DNA sequence in its linear order. Given the advent of electrophoresis (see Milestone 1), the questions became how could one generate fragments at every position and how could the terminal base of the fragments be distinguished?

Three methods were revolutionary in achieving these goals. The first, cleaving single-stranded DNA (ssDNA) or denatured double-stranded DNA (dsDNA) labelled at one end, was published by Maxam and Gilbert in 1977. Fragments were generated in two steps: the base was removed, and then the weakened sugar bond was broken by the addition of alkali or amines. In this chemical method, the four bases were not specifically determined; rather, one of the four reactions cleaved at pyrimidines, one cleaved at C, and the other two had a preference for cleaving G > A or A > G. The four pools of fragments were separated in different lanes of a polyacrylamide gel and the sequence, with a read length of ~100 bases, was deduced from the ladder of bands.

Later that year, Sanger and colleagues published an enzymatic sequencing protocol. This method did not involve DNA breakage, but exploited the fact that dideoxynucleotides (ddNTPs), lacking a 3′-hydroxyl, cannot be extended by DNA polymerase. Consequently, one could set up four DNA synthesis reactions containing the same ssDNA template and primer, DNA polymerase, a mixture of the deoxynucleotide (dNTP) and ddNTP forms of one of the nucleotides, and the remaining three dNTPs (one of which was labelled). As both the ddNTP and dNTP were present, DNA polymerase would sometimes incorporate the correct dNTP, allowing further polymerization, and at other times would incorporate the ddNTP, causing chain termination. After DNA denaturation and polyacrylamide gel electrophoresis, a series of labelled bands corresponding to termination at a specific nucleotide could be read.

It was appreciated that in order to sequence genomes, it would be necessary to automate sequencing and gather information in real time. Into this void stepped the Hood laboratory with a third technique, a variation of the Sanger method that bypassed the rate-limiting step — detection of the labelled bands by autoradiography. Rather than using labelled dNTPs, the primer oligonucleotides in the four reactions were attached to fluorophores with different emission maxima. The four reactions yielded fragments tagged with different fluorophores, so that they could be combined and run in a single column gel. Simultaneously, the group developed a detection system that used a laser to read the fluorophore signals, and an analytical program that resolved the raw data into a series of peaks that corresponded to the sequence. When a single lane contained all four reactions, 200 bases were read with ease. Subsequently the fluorophores were attached to ddNTP terminators, removing the need for tagged primers and allowing all four reactions to be performed in one reaction, slab gels were replaced with capillaries and read lengths were extended to 600 or more bases.

These approaches facilitated the explosion of sequence-gathering studies that have evolved into our current bioinformatics-driven research.

Angela K. Eggleston, Senior Editor, Nature

ORIGINAL RESEARCH PAPERS Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA* **74**, 560–564 (1977) | Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977) | Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986)

WEB SITE
DNA sequencing from Wikipedia: http://en.wikipedia.org/wiki/DNA_sequencing