

This Week's Citation Classic

Sutcliffe J. G. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp.* 43:77-90, 1979.
[Biological Laboratories, Harvard University, Cambridge, MA]

This paper describes the nucleotide sequence of pBR322 and the strategy that was developed to ensure the accuracy of the results. The biological features encoded by the plasmid are interpreted through the sequence, as are features that enhance its value as a cloning vector. [The SCI® indicates that this paper has been cited in more than 1,600 publications, making it the most-cited paper from this journal.]

The Sequence of pBR322

J. Gregor Sutcliffe
Department of Molecular Biology
Research Institute of Scripps Clinic
10666 North Torrey Pines Road
La Jolla, CA 92037*

The middle 1970s were an exciting time to be a graduate student in molecular biology, especially if one were fortunate enough to be in the large group jointly led by Jim Watson, Wally Gilbert (my thesis adviser), Klaus Weber, and David Dressler at the Harvard Biological Laboratories. Many important papers were written during that period by members of the group, including one by Gilbert and Allan Maxam describing a chemical method for determining DNA sequences.¹

DNA sequencing technology was crude at that time, and previous methods were not highly reliable. Indeed, the landmark sequence of the ϕ X174 genome from Sanger's group² using DNA analysis combined with brute force isolation and sequencing of all of the ϕ X174 RNA transcripts and protein products had just been published, concluding that "errors and uncertainties can only be eliminated by more laborious experiments and...it would probably be a long time before the complete sequence could be established. We are not certain that there is any scientific justification for establishing every detail...." The Maxam-Gilbert method looked capable of producing more accurate data, but it had been applied only to relatively short regions of DNA.

Gilbert and his colleague, Jeremy Knowles, wondered whether the sequence of an entire protein, β -lactamase, could be obtained solely from structural analysis of its gene. Partial amino acid sequence data, which Knowles kept locked in his desk, were available from R.P. Ambler and G.K. Scott in Edinburgh. I was chosen for this test of the

Maxam-Gilbert sequencing method in part because I had lobbied for an assessment of its accuracy. I selected the β -lactamase gene of pBR322 for study because it was evident this was a superior cloning vector and any information about it would be valuable. Over seven months, I learned the method, developed strategies for studying an entire gene, and scaled-up the technology, especially gel electrophoresis, since then it could require 10 gels to obtain 100 bases (now it's kilobases/gel). Every base was critical as errors could alter the apparent triplet reading frame and lead to an imaginary predicted protein sequence. Thus I developed careful checking procedures to give me confidence in the final 1,100 base β -lactamase gene sequence.³ The moment of truth came over tea at Knowles's house on a weekend afternoon. We went residue by residue, comparing my handwritten nucleotide and deduced amino acid sequence to the Ambler-Scott data. We found only one discrepancy that turned out to result from a strain difference rather than error.

The comparison was exhilarating and its result gratifying given the effort that went into achieving it. This increased my incentive to complete the entire pBR322 sequence, a project that required more than 400 gels and was finished six months later. The final 4,362 base sequence was proofread with the help of most members of the group. Over the years, only one error has been found, although an embarrassing one, as it affected the interpretation of the tetracycline resistance mechanism. Nevertheless, the magnitude of the project was unprecedented and certainly changed the perception I, and probably others, had of the power of DNA sequence analysis.

This and a related paper containing restriction cleavage maps derived from the sequence⁴ have been cited because the data they contain have direct application in many studies. pBR322 became the cloning vector of choice, in part because of the availability of its sequence. However, the greater value of the work was that it demonstrated the startling accuracy and speed with which kilobases of DNA sequence could be determined. This realization has contributed to the consideration of the sequence of the human genome as an achievable goal. That carefully determined sequences provide solid foundations for experimental work is evident in most molecular approaches taken in the 1980s.

1. Maxam A M & Gilbert W. A new method for sequencing DNA. *Proc. Nat. Acad. Sci. USA* 74:560-4, 1977. (Cited 5,555 times.)
2. Sanger F, Air G M, Barrell B G, Brown N L, Coulson A R, Fiddes J C, Hutchison C A, Slocombe P M & Smith M. Nucleotide sequence of bacteriophage ϕ X174. *Nature* 265:687-95, 1977. (Cited 880 times.)
3. Sutcliffe J G. Nucleotide sequence of the ampicillin resistance gene of *E. coli* plasmid pBR322. *Proc. Nat. Acad. Sci. USA* 75:3737-41, 1978. (Cited 405 times.)
4. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. *Nucl. Acid. Res.* 5:2721-8, 1978. (Cited 780 times.)

*Received August 3, 1990

1-22