From the pages and gels of history . . .

Consolidating DNA Sequencing

In his first article which appeared in the April 2001 issue of the Australian Biochemist, Geoff Grigg recounted the story of the early days of DNA sequencing in Fred Sanger's laboratory in Cambridge, up to 1973. In this, the second and concluding article, Geoff goes on to describe the refinement of the dideoxy method of DNA sequencing, the foundation of genomic analysis today. he also describes the development of methods for sequencing 5-methyl cytosine in genomic DNA.

I was back in Cambridge three years later. During these three years Fred Sanger had devised the + and – methods of sequencing DNA and had started to get results with the dideoxy method. The dideoxy method was the more powerful, and the + and – methods had been dropped and forgotten, so I shall not resurrect their memory.

The only hold-up to exploiting the dideoxy method was the availability of the chain-terminating dideoxyribonucleotide triphosphate reagents. Fred had synthesised the first one he used, dideoxythymidine triphosphate, only after a bit of a struggle.

When I visited him and enquired about progress he offered to give me a sample to try back in my Sydney lab. He pulled out a 2.5 cm diameter tube with a greasy looking smudge down one side, scraped off a generous portion of the smudge into a second tube which he labelled and presented to me. "How much is there?" I enquired. "I don't know," he replied, "there was not enough in the prep to weigh it. However, it works".

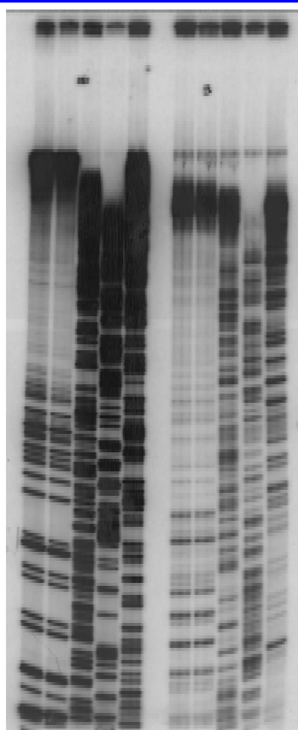
The four dideoxy triphosphates, incorporated one at a time into four different polymerase reaction mixes as a minor component, terminate synthesis in a nascent growing oligonucleotide chain when a dideoxy compound is incorporated. One of his early, unpublished, sequencing-gel autoradiographs is reproduced below.

Even by 1978 when I next visited the Sanger lab, Bart Barrell, Fred's erstwhile technician and Alan Coulson his current one, claimed to be able to sequence 40,000 residues a day each, using manual methods. Only a few years before, 50 residues a year was the norm.

Also in 1976, Fred's lab adopted the shotgun method of sequencing based on random breakage of DNA, usually by sonication, followed by cloning into a single stranded DNA phage to give a suitable quantity of each fragment for sequencing. The sequenced fragments were then assembled by computer. The dideoxy method is the basis of all the main DNA sequencing efforts around the world today.

The currently used protocol in the Sanger lab was printed as a 40 page booklet and handed out to new arrivals in the lab. This process seems virtually the same as that 'reinvented' by the large genomics company, Celera Inc., some 20 years later.

My visit in 1976 to Cambridge when Fred Sanger had just got the dideoxy method to work was enlivened by a week-end cruising the coast of East Anglia in Fred's new yacht which he had built in his backyard over the previous few years. I had introduced him to the pleasures of sailing on Sydney Harbour in 1967 and 1969 and he had fallen in love with the pastime.



Gel rescued by Geoff Grigg on the day Sanger retired from LMB, Cambridge.

Three of us were together on the yacht - Fred, Cesar Milstein and me. The previous year Milstein had just published, with Georges Kohler, his monoclonal antibody paper. I remember quizzing him on whether they had patented the method and its uses since it seemed to have important applications in a number of areas. "Of course I had thought of doing so," was Cesar's reply, "however the British Technology Group, which the MRC had consulted, said monoclonal antibodies had no commercial value so a patent was not applied for". Foolish advice!

Fred Sanger did not patent any of his DNA sequencing methods. He strongly objected to the patenting of methods whose uses, he thought, should be free of any constraints and available for all to use. The dideoxy method was to earn Sanger a second Nobel Prize and an OM to add to his CBE and CH. He shared the Nobel Prize with Walter Gilbert, who with Maxam had devised another DNA sequencing method which was used in the early 1980's particularly in the USA.

The Maxam and Gilbert (chemical) method involved treating equal samples of a single stranded DNA sample which had been radioactivity labelled at its 5' end in turn, with each of a number of reagents which selectively attacked the sugar-phosphate backbone of the DNA at sites adjacent to particular nucleotides. Doses of the reagents were controlled so as only to cleave a small proportion of available sites.

The treated DNA, which was now fragmented, was then size-fractionated by electrophoresis on a thin polyacrylamide gel. By comparing the pattern of bands representing the oligonucleotides in the several different base-specific treatment groups the base sequence of the target DNA strand could be deduced. Initially this chemical method seemed to be preferred, at least on the western side of the Atlantic, over the Sanger method; however, its usefulness was restricted by the

From the pages and gels of history (contin.) . . .

Consolidating DNA Sequencing

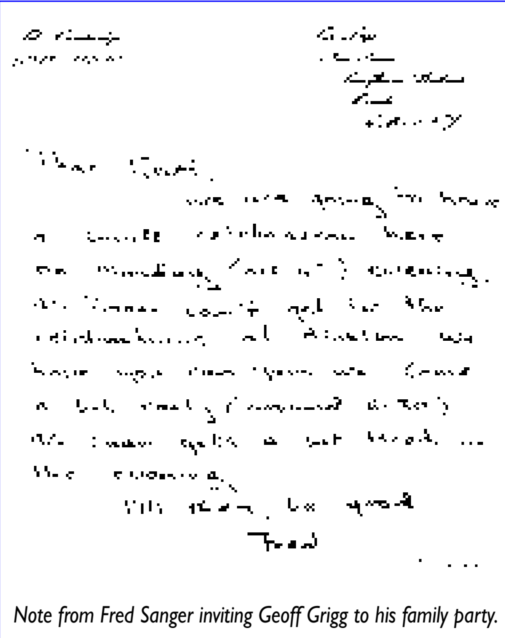
need first to isolate as a pure sample a DNA fragment containing the target sequence. The greater power and speed of the dideoxy method finally determined its choice by most laboratories.

The simplicity of the dideoxy method allowed the development of robotic methods to automate the large scale sequence analysis of various genomes ranging in size from viral pathogens to the human genome. The use of radioactive markers to visualise the DNA copies was replaced by fluorescent markers later. Because these came in various colours the positions of each of the four bases A, T, G and C could be determined from the fractionation of mixtures of four polymerase reactions.

An important factor in the useful exploitation of these new methods was the establishment around the world of both publicly and privately funded laboratories dedicated to sequencing the DNA of these genomes. Of these one of the more important was the new Sanger Centre and an associated Informatics research laboratory, established with the aid of generous funding provided by the Wellcome Foundation of the United Kingdom. It was operated by the Wellcome Foundation, the Medical Research Centre of the UK and, for the Informatics laboratory, the European Molecular Biology Organisation. Dr Bridget Ogilvie, Director of the Wellcome Foundation, had been approached informally by Sir Aaron Klug, then Director of the Laboratory of Molecular Biology at Cambridge, to enquire of their interest in supporting a major DNA sequencing initiative.

The Foundation made a rapid decision to throw major resources into the project straight away to support a DNA sequencing centre with John Sulston as Director. Within a very short time a large site, Hinxton Hall, with an existing laboratory building on it at the village of Hinxton a few miles south of the MRC laboratory had been acquired, and the first scientists had commenced work.

At the small opening ceremony, attended by a few MRC, Wellcome Foundation staff and a handful of outsiders, the Chairman of the Wellcome Foundation spoke of giving the Americans a run for their money, and Fred Sanger formally opened the Centre. A rather larger party to which all the LMB people were invited in the evening followed at the site's canteen. Fred Sanger's wife had just been discharged from hospital so he chose to attend his own small family party at home organised by his daughter and nephew rather than the big party event, where champagne flowed freely.



Note from Fred Sanger inviting Geoff Grigg to his family party.

Within a little over a year a series of large new laboratories had been designed to house some 600 staff scientists, and building work started at Hinxton Village. The original laboratory on the site was demolished and the handsome 18th century buildings of the Hinxton Hall estate were faithfully restored. Subsequently Wellcome Foundation acquired most of the village and a further 2000 acres for future use. Other large sequencing laboratories, mainly in the USA and France, were also now well established and into the race to be first to sequence the genomes of this, that, and the other organism, and as much of the human one that they could. This latter task is now largely completed.

The *next* large sequencing task of the Sanger Centre, as recently announced, is to analyse the human genome for its complement of (tissue specific) methylcytosine residues, because of the importance of 5-methylcytosines as signals in the genome in normal and abnormal development and gene function.

Amongst the principal weapons the Centre will use for this task is the bisulphite method of sequencing 5-methylcytosine in genomic DNA. But before discussing this method which was invented in our laboratory, let us go back a half century to Hotchkiss's discovery of 5-methylcytosine as a relatively common fifth base in DNA. Twenty years on it proved to be important in explaining the restriction-modification phenomenon in various bacterial species.

Sanger Centre building at Hixton, Cambridgeshire, UK (reproduced with permission from the Wellcome Trust Medical Photolibrary).

Consolidating DNA Sequencing

Sequencing 5-methylcytosine (5-meC)

In 1975 Holliday and Pugh in England and Arthur Riggs in the USA simultaneously, and quite coincidentally, suggested a role for 5-meC as a regulatory signal in both normal and pathological development of higher plants and animals. They also suggested how, once established, a methylation pattern of cytosines in DNA might be maintained without change from a parent DNA to daughter copies.

This theory has been confirmed in a number of ways. Moreover, the role of cytosine methylation and its importance in the regulation of genome expression, in both normal and pathological development in higher plants and animals, X-inactivation and imprinting, is now well established. Local scientists, including those at the North Ryde and Canberra campuses of CSIRO, including Peter Molloy, Robin Holliday (who had moved to Australia), Liz Dennis, Jean Finnegan, Jim Peacock and their staff, as well as Sue Clark, Doug Millar and their associates at the Kanematsu Institute in Sydney, played a significant part.

The early methods of sequencing 5-meC in DNA relied on sequence selective breakage of DNA either by enzymes such as certain restriction enzymes, or by chemical reactions. These methods had some limitations; application of the restriction enzyme method was limited to those cytosines (ca. 10%) which were part of the specific sequence recognised by an enzyme. Most of the methods required the use of significant amounts of genomic DNA.

About 10 years ago a better method was invented. This 'bisulphite' method was of general applicability and required very little DNA - that from a few cells was all that was necessary. Moreover it gave more information than any existing method. It was developed in the North Ryde labs of CSIRO's Laboratory of Molecular Biology, by groups from CSIRO and the Kanematsu Institute.

Over the years it has been steadily improved to make it faster and more user-friendly. It is now in general use worldwide and is the method being employed by the consortium of large European labs, led by the Sanger Centre at Hinxton near Cambridge, in their recently announced project to sequence the methylcytosine content of large stretches of the human genome.

Recently a number of features of the bisulphite procedure have been adapted by various labs, including ours, to more rapidly derive information on the methylation state of short key regions of the genome such as gene promoters. The new method devised by us will be published later this year.

The basic idea of the bisulphite method was conceived in 1988. It had seemed to me that a sequencing method which relied for its effectiveness on selective modification of cytosines versus methylcytosines, rather than on selective breakage of genomic DNA at cytosine methylation sites, might have advantages because it would need less DNA and it should be of more general applicability.

It had been known for some 20-30 years that a number of reagents reacted with cytosine but not with methylcytosine. They formed adducts across the 5-6 double bond of cytosine, and by so doing destabilised the NH_2 amino group of the cytosine ring. The net result of such reactions carried out in aqueous solution was to convert cytosine adducts to uracil adducts. Since uracil on an oligonucleotide template behaves like a thymine in a polymerase reaction, the net result of such a reaction followed by a PCR step would be that all the cytosines would be converted to thymines, whereas all the methylcytosines would be represented by cytosines. This process was the basis of the method.

After discussing the chemistry of cytosine modification with Dan Brown in Cambridge when I was there in 1988, I decided to use bisulphite as the key reagent to distinguish between cytosine and 5-methylcytosine. The chemistry of the reaction between the bases and bisulphite had been worked out by Hikoya Hayatsu in Japan in the 1970's.

One consequence of the change of cytosine bases to uracil/thymine bases in DNA would be that the erstwhile complementary strands of the double helix no longer would be complementary. So, if desired, each could be copied independently of the other by the choice of appropriate primer oligonucleotides. This principle enables certain previously untestable presumptions to be checked.

These ideas were reduced to practice by Marianne Frommer and her group at the Kanematsu, aided by various resources from others in the CSIRO laboratory where they worked. Their results from a sequencing reaction using the bisulfite method gave a positive readout – each band in the C slot of a fractionation gel represented the position of a 5-meC residue – so they were easy to see.

The first paper on the bisulphite method appeared in *Proc. Natl. Acad. Sci. USA* in 1992; and various improvements, mainly from the Kanematsu and our labs, emerged later. It is now in general use worldwide.

Methodology is never a static process – and despite the power and efficiency of the dideoxy method, and its derivative methods which include the bisulphite method for methylcytosine residues in DNA, other ideas of sequencing DNA such as those using mass spectroscopic techniques continue to be developed. No matter how effective they may prove to be, they cannot diminish the revolutionary effects of the dideoxy method on biological theory and practice over the past 20 years – all because of the bright ideas of a few people.

Geoff Grigg can be contacted at the CSIRO Division of Molecular Sciences, PO Box 184, North Ryde NSW 2113