learn the techniques. The rest of this story is better known. Many hundreds of specifically targeted mouse mutations have been made and the technique, although still not trivial, may now merit no more than a few lines' mention in experimental genetics papers. Almost any specific genetic change may now be generated, selected and verified in culture before being converted to the germ lines of mice, and this is the experimental genetics that is illuminating our understanding of the mammalian genome physiology and human function in health and disease.

I set out to derive a 'tractable' system for following mRNA changes coincident with embryonic cell differentiation. ES cells now provide the culture system and, at long last, methods for genome-wide monitoring of mRNA have come of age in cDNA microarrays. I am now putting the two techniques together, and results are beginning to emerge from this work (Fig. 2).

Forty years with homologous recombination

OLIVER SMITHIES

Toolmakers—and I suspect that the three of us being honored by the Lasker Foundation fit into this category—are fortunate people. They see problems, invent tools to solve them and enjoy the solutions, which often demonstrate new principles that were not part of the original thought. As a bonus, they also enjoy the vicarious pleasure of seeing other people use the same tools to solve very different problems. Yet the invention of an effective scientific tool is rarely an isolated event; there are often many prior experiences that trigger the inventive thought, and there may be various unexpected additional problems to solve before the toolmaker can bring a nascent idea into practice.

The chain of events leading to my contributions to the use of homologous recombination to modify genes in the mouse genome began over 40 years ago as an unplanned consequence of my somewhat serendipitous invention in the 1950s of an earlier tool—high-resolution gel electrophoresis—to solve a completely non-genetic problem. On 26 October 1954, during final pre-publication tests of my starch-gel electrophoresis system (the immediate forerunner of one of molecular biologists' primary tools, polyacrylamide gel electrophoresis), I ran a sample of serum from a female. My notebook (Fig. 1) has the entry that the pattern was "Most odd—many extra components." For about a week I enjoyed the misconception that I had discovered a new way of telling males from females. But this 'sexy' hypothesis soon gave way to the idea that "hereditary factors may determine the serum groups"[19] and, with the help of Norma Ford Walker, who began my education as a geneticist, this was soon proven correct[20]. The field of normal human protein polymorphic variants was seeded!

The hereditary variations we had discovered proved to be in the hemoglobin-binding serum protein haptoglobin, and their details were worked out during a happy collaboration between George E. Connell, Gordon H. Dixon and me in the early 1960s. The haptoglobin alleles *Hp1F* (fast) and *Hp1S* (slow) encoded polypeptides differing by two amino acids, but the third allele, *Hp2,* seemed to be a tandem joining together of sequences from *Hp1F* with sequences from *Hp1S*. The then-chairman of my department at the University of Wisconsin, James F. Crow, on being asked how the *Hp2* allele might have arisen, directed me to the *Bar* locus in *Drosophila* with its fascinating history of repeated 'mutations' resulting from unequal crossing over[21]. This led us to hypothesize that the *Hp2* allele was formed by a unique non-homologous recombinational event that joined the end of *Hp1F* to the beginning of *Hp1S* (ref. 22). *Hp2* therefore contained a small intragenic tandem duplication. The Bar gene in *Drosophila* is also a unique tandem duplication, but it is large enough to be visible when the fly salivary chromosomes are under the microscope. Yet the consequences of the tandemly repeated sequences in *Bar* and in *Hp* are completely comparable. In both cases, subsequent predictable unequal homologous crossing over events occur, which generate a new triplicate product and regenerate the singleton: B–B × B–B leads to B–B–B + B.

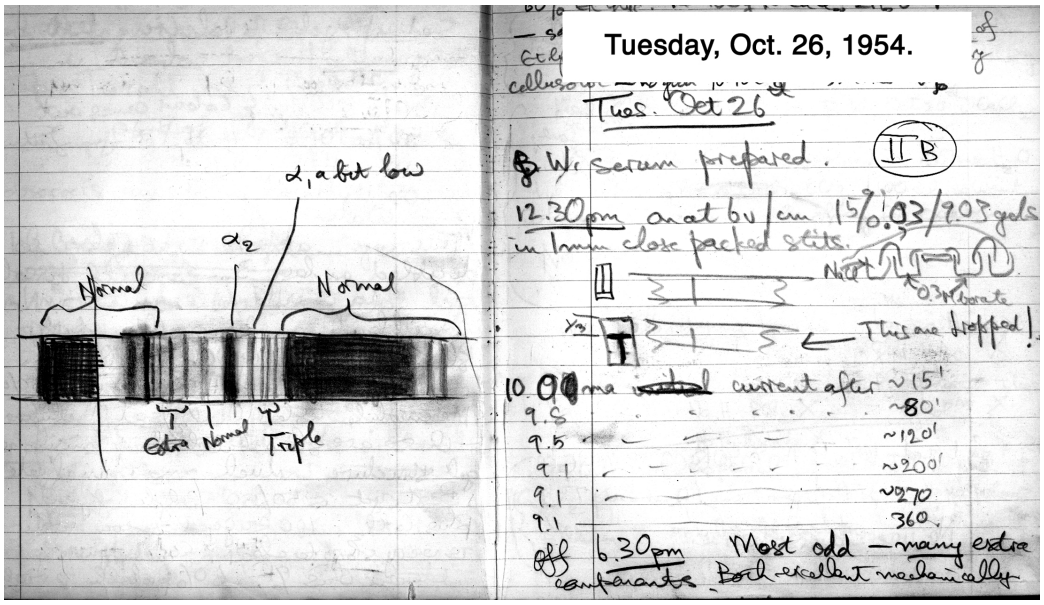

Fig. 1 Pages 97 & 98 from Smithies' lab notebook "Physical IV", 1954.

I found the predictability of homologous recombination seductive, and enjoyed enormously hypothesizing that antibody variability might be achieved by homologous recombination between tandemly arranged sequences[23]. The hypothesis turned out to be incorrect in mammals, but was remarkably close to being correct in chickens. Homologous recombination reappeared in my experimental science in the early days of cloning human genes when we were determining the nucleotide sequences of the two

human fetal globin genes, $^{G}\gamma$ and $^{A}\gamma$. The nucleotide sequence data showed distinct evidence that an exchange of DNA sequences had occurred between these two genes, as a result of homologous recombination in the form of a gene-conversion event[24].

With 20 years of experimental and theoretical exposure to the ubiquity and predictability of homologous recombination in the human genome when suitable sequences were present, and an inexhaustible supply of normal β-globin DNA now available by cloning, a not-unexpected thought kept coming into my mind, namely that correction of the sickle-cell β-globin gene mutation should be possible by homologous recombination between 'corrective' DNA introduced into a cell and a mutated gene present in the cellular genome. However, I had no way of estimating the likely frequency of such an event, which I suspected might be low, nor could I think of any easy way of detecting the event if it occurred. Then in the spring of 1982, I reviewed for my genetics class a paper published in *Nature* 1 April 1982 (ref. 25). The authors of this paper were searching for a mutated gene in a carcinoma cell genome that changes normal cells into focus-forming transformants. Their strategy for isolating this gene was elegant, although complex. But I realized that it could be adapted to determining whether homologous recombination can place corrective DNA into a chosen place in the genome.

I have lost the scraps of paper on which this realization was assembled into a plan, but on 22 April 1982, 3 weeks after the Goldfarb paper was published, I wrote in my lab notebook (Fig. 2) an outline of how I thought a test of corrective gene targeting might be made. In my more than 100 lab notebooks, this page is the one I like best! The principle of the proposed "assay for gene placement" is simple, although its execution was not. It depends on detecting in the genomes of correctly targeted cells the bringing together of DNA sequences present in the targeting construct but not in the genome and DNA sequences that are in the target locus but not in the incoming DNA. Three levels of selection were in the assay. I fondly imagined (probably incorrectly) that the power of my assay was therefore about 10^5 (for the thymidine kinase selection in eukaryotic cells) $\times$ 10^5 (for the *sup F* selection in prokaryotic cells) $\times$ 10^6 (for the β-specific probe hybridization). Even if targeting were random, I should be able to detect it!

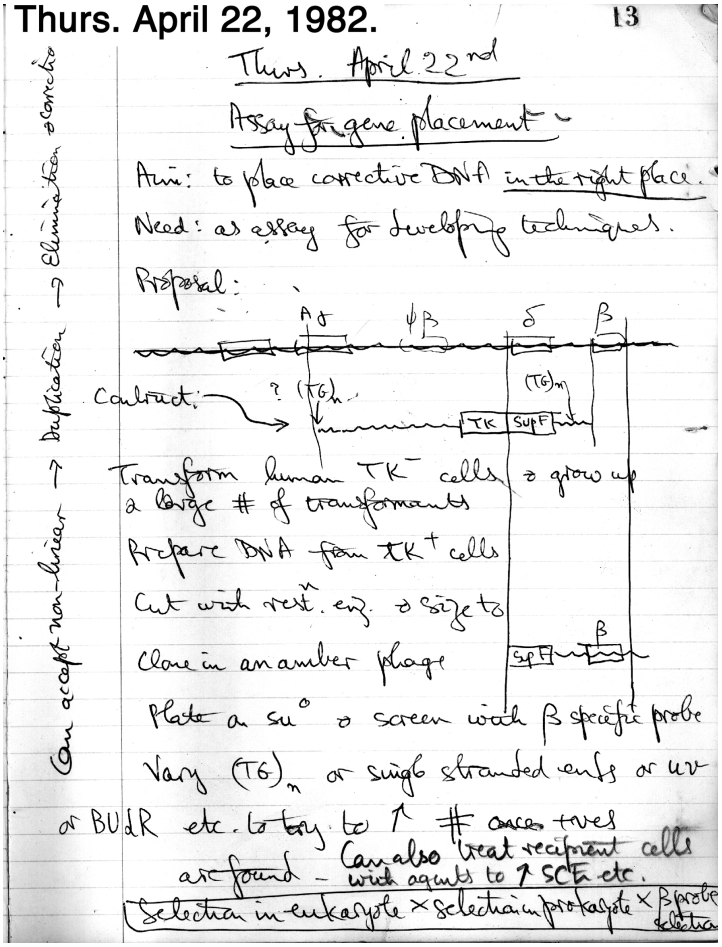

Fig. 2 Page 13 from Smithies' lab notebook γ, 1982.

It took 3 years and the invaluable help of my postdoctoral fellow Ron Gregg, my visiting professor Sallie Boggs, my technician Mike Koralewski and my longtime collaborator Raju Kucherlapati to go from this notebook page to successful proof that homologous recombination can be used to target a chosen gene in a mammalian cell. We first tested the scheme using human bladder carcinoma cells and calcium phosphate–DNA precipitates to introduce the DNA into the cells. We obtained no positive signals with these cells. My graduate student Karen Lyons pointed out that bladder cells might not express a selectable marker when it is introduced into the β-globin locus, as they do not express β-globin. So I began again with modified mouse erythroleukemia cells that do express the β-globin gene. But these cells only grow in suspension and cannot be transformed with calcium phosphate, and electroporators were not yet available commercially. So I designed and built my own, with a plastic bath tub, part of a test tube rack, a glass plate and some silicon-controlled rectifiers (Fig. 3). We used this apparatus for all the key experiments.

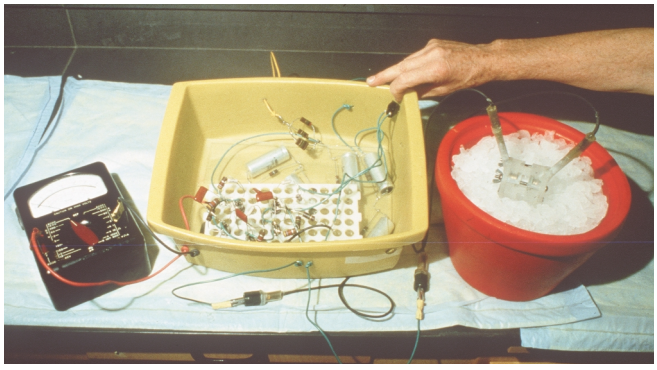

Fig. 3 Electroporator for introducing exogenous DNA into target cells, Smithies' lab notebook ι, 1984.

The first real, albeit indirect, evidence that the experiment was working was obtained on 30 January 1985, the happy day when we first detected bacteriophages that grew because they had picked up the *supF* gene and that also hybridized to the β-globin probe. We were now fairly certain that planned modification by homologous recombination had been achieved. To obtain direct evidence, over the next 3 months we did a 'sib-selection' procedure (bacteriophage assays on total DNA from decreasingly small pools of cells) until we were down to three 20-colony pools, one of which was still positive by the

bacteriophage assay. Individual colonies from this pool were tested on 18 May 1985 by Southern blot analysis (a direct assay). DNA from one of the colonies produced a hybridizing fragment of the correct size (Fig. 4), and we were 'home'!

I presented the results of our work at a Gordon Conference in 1985, and told the attendees the true story that, as I developed the critical gel autoradiograph, which we knew would provide the first direct test of whether or not the target gene had been modified, I was thinking that we had been a long time (3 years) knowing that our experiment was working only by indirect evidence—much like being an airplane pilot on instruments in the clouds. The autoradiograph was the moment of truth, comparable to the moment when you descend below the clouds and no longer depend on the indirect indications of your instruments: The runway is either there or it is not! The thrill of seeing it never pales. For the remainder of that meeting, other investigators would say, as they pointed to a desired result, "And there is my runway!" We published our results in the 19 September 1985 issue of *Nature*[18].

Nonetheless, our 'runway' was exceedingly difficult to find. In only about one in a million treated cells was homologous recombination achieved. Such a low frequency of gene targeting was not much use for gene therapy. And the assay, like my doctoral-thesis method of measuring osmotic pressures[26], was remarkably good at doing what it was designed to do, but both methods were impossibly laborious. No one, not even me, ever used either again. So, what to do? The first order of business was to try to improve the method. For this we needed an easier target, preferably one whose targeting could be assessed directly. The hypoxanthine phosphoribosyltransferase gene *(HPRT)* was an obvious choice, and so Ron Gregg began a series of attempts to correct a mutated *HPRT* or to inactivate a wild-type copy of the gene using homologous recombination.

We also needed to replace the bacteriophage recombinant fragment assay with something easier. Kary Mullis' new PCR tool could in principle detect recombinants. We could choose one primer specific to the incoming DNA and another primer specific to the target gene. PCR amplification would then only occur when the two primer sequences were juxtaposed by the desired homologous recombination. But, again, there were no commercial PCR machines available. So we made our own out of three old water-baths, home-made controllers and hot water valves used in domestic heating systems[27]. We still use it! Its six hoses look like octopus arms; for obvious reasons we call it 'hexapus'. The ease of this PCR-based recombinant fragment assay made screening for homologous recombinants much less difficult.

Meanwhile, at a 1985 Gordon Conference, I heard Erwin Wagner talk about Martin Evans' embryonic stem cells[12], which after injection into blastocysts can produce living progeny mice. Here was a more promising use of our one-in-a-million targeting skill. We could generate planned mutations or correct existing mutations in tissue culture, even if it took millions of cells, and expect to transfer the alterations into living mice. A visit to Erwin Wagner led to my contacting Martin Evans who, with typical generosity of spirit, personally brought some of his EK CC-1 cells to us in November 1985. My plan was "to use these to get *HPRT*$^-$ by recombination and get chimeras or germline by blastocyst route." Martin also put me in touch with Tom Doetschman, an American postdoctoral fellow wanting to return to the United States, who had personally isolated embryonic stem cells (now called ES cells) while in Rolf Kemmler's laboratory. He joined our group in late 1986.

At this point, Nobuyo Maeda and I attended a conference in Scotland at which Martin Evans and Martin Hooper both reported that they had obtained *HPRT*$^-$ mutant ES cells in tissue culture experiments. Nobuyo recognized that, in the course of helping Ron Gregg, she had already made a construct that could correct either of their *HPRT*$^-$ mutant cells. We told Evans and Hooper about this, and both immediately agreed to collaborate with us: Martin Hooper sent his mutant cells (TG-2a) to us, and we sent our construct to Martin Evans. Tom Doetschman tried Nobuyo's construct on the TG-2a cells. The very first experiment worked, and we published our results in *Nature* 10 December 1987 (ref. 28). But it still took 2 more years of valiant effort, spearheaded by my postdoctoral fellow Bev Koller, to accomplish the mouse blastocyst injections leading to chimeras, to obtain progeny with the altered gene and to report the "Germ-line transmission of a planned alteration made by homologous recombination in embryonic stem cells."[29]

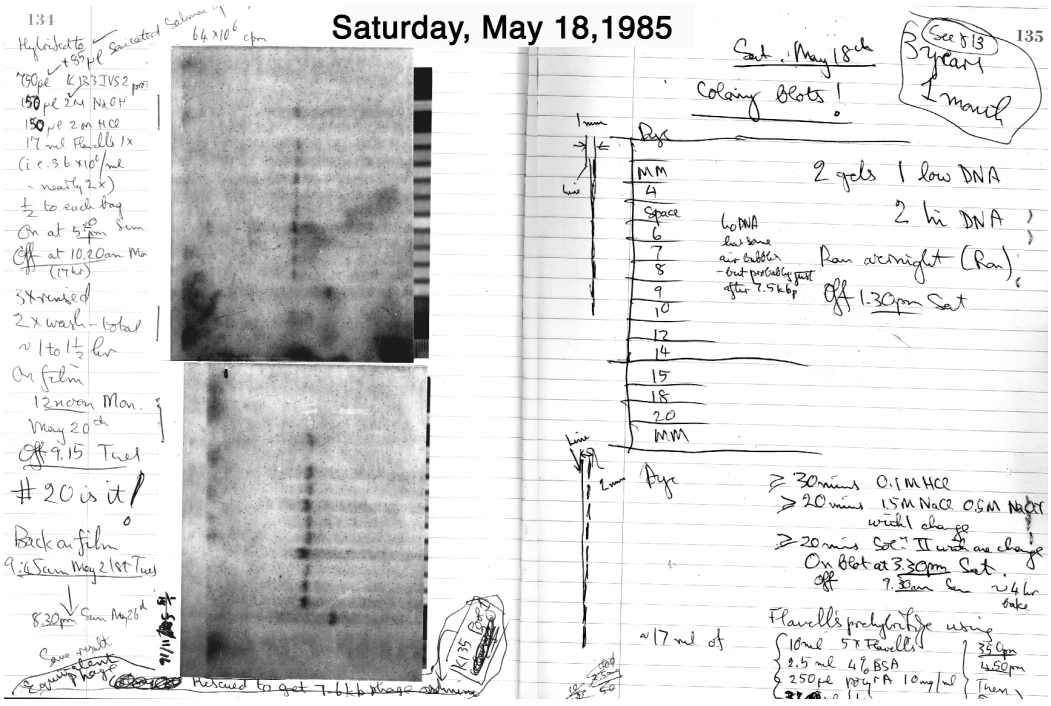

Fig. 4 Pages 134 & 135 from Smithies' lab notebook κ, 1985.

We next turned our attention to problems related to human diseases, beginning by

replicating the infrequent (1 in 50,000) but simply inherited human genetic disease cystic fibrosis, and some of the hemoglobinopathies (1 in 100–1,000), and progressing to using the method to decipher the genetic complexities of much more common but also much more complex conditions, such as atherosclerosis (1 in 2) and hypertension (1 in 5)[30]. We had an important conceptual change early in our studies of the genetic complexities of hypertension. This was a shift from considering absence of gene function as a principal cause of disease (as is the case with the uncommon disease cystic fibrosis) to considering the possibility that inherited quantitative variations (perhaps even normal variations) in gene expression might be more important in causing the complex common conditions. To investigate this possibility experimentally, we devised a 'gene-titration' method, in which two complementary forms of homologous recombination are used to vary the number of copies of a candidate gene from one through four[31]. The 'one-copy' animals are heterozygous for a wild-type allele and a deleted copy. The 'three-copy' and 'four-copy' animals use a complete tandem gene duplication reminiscent of *Hp2*. In the most dramatic of these experiments, the resulting gene expression varies linearly with copy number from ½× to 2× normal. With the current emphasis on single-nucleotide polymorphisms and functional genomics, it is likely that mice obtained by homologous recombination will prove to be of great value in establishing whether a genotype associated in humans with a complex phenotype could in fact cause the condition.

Obviously I continue to enjoy using the tool for which we are being honored to solve problems of interest to me. And when I open any current issue of the main journals covering biological science, I am very likely to have the vicarious enjoyment of seeing some other investigators' use of homologous recombination to modify their chosen gene in the mouse genome.

Generating mice with targeted mutations

MARIO R. CAPECCHI

Mutational analysis is one of the most informative approaches available for the study of complex biological processes. It has been particularly successful in the analysis of the biology of bacteria, yeast, the nematode worm *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*. Extension of this approach to the mouse, though informative, was far less successful relative to what has been achieved with these simpler model organisms. This is because it is not numerically practical in mice to use random mutagenesis to isolate mutations that affect a specified biological process of interest. Nonetheless, biological phenomena such as a sophisticated immune response, cancer, vascular disease or higher-order cognitive function, to mention just a few, must be analyzed in organisms that show such phenomena, and for this reason geneticists and other researchers have turned to the mouse. Gene targeting, the means for creating mice with designed mutations in almost any gene[32], was developed as an alternative to the impractical use of random mutagenesis for pursuing genetic analysis in the mouse. Now gene targeting has advanced the genomic manipulations possible in mice to a level that can be matched only in far simpler organisms such as bacteria and yeast.

The development of gene targeting in mice required the solution to two problems: How to produce a specific mutation in a chosen gene in cultured mammalian cells, and how to transfer this mutation to the mouse germ line. Oliver Smithies' laboratory and mine worked independently on solutions to the first problem. Martin Evans' laboratory provided the basis for a solution to the second problem.

Early experiments

Our entry into what became the field of gene targeting began in 1977. At that time, I was attempting to improve the efficiency with which new genes could be introduced into mammalian cells. It had just been demonstrated by Wigler and Axel that cultured mammalian cells deficient in thymidine kinase (Tk^-) could be transformed to Tk^+ status by the introduction of a functional copy of the herpes thymidine kinase gene (*HSV-tk*)[33]. Although an important advance for the field of somatic cell genetics, their protocol—the use of calcium phosphate co-precipitation to introduce the DNA into cultured cells—was not efficient. With this method, incorporation of functional copies of *tk* occurred in only one per million cells exposed to the DNA–calcium phosphate co-precipitate. Using a similar selection scheme, I sought to determine whether I could introduce a functional *tk* into Tk^- cells using very fine glass needles to inject DNA directly into nuclei[34]. This procedure proved extremely efficient. One cell in three that received the DNA stably passed the functional *tk* to its daughter cells. The high efficiency of DNA transfer by microinjection made it practical for investigators to generate transgenic mice containing random insertions of exogenous DNA. This was accomplished by injection of the desired DNA into nuclei of one-cell zygotes and allowing these embryos to come to term after surgical transfer to foster mothers[35–39].

Efficient functional transfer of *HSV-tk* into cells required that the injected *tk* be linked to other short viral DNA sequences[34]. It seemed plausible that highly evolved viral genomes might contain bits of DNA that enhance their ability to establish themselves within mammalian cell genomes. I searched the genome of the lytic simian virus SV40 for the presence of such sequences and found one near the origin of viral DNA replication. When linked to *HSV-tk,* it increased the transforming capacity of the injected *tk* by 100-fold. I showed that the enhancement did not seem to result from independent replication of the injected *HSV-tk* DNA as an extra-chromosomal plasmid, but rather that the efficiency-enhancing sequence was either

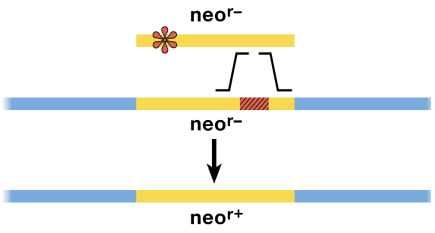

Fig. 1 Regeneration of a functional *neo*^{r} by gene targeting. The recipient cell contains a defective *neo*^{r} with a deletion mutation (*). The targeting vector contains a 5′ point mutation. With a frequency of 1 in 1,000 cells receiving an injection, the deletion mutation in the chromosomal copy of *neo*^{r} is corrected with information supplied by the targeting vector.