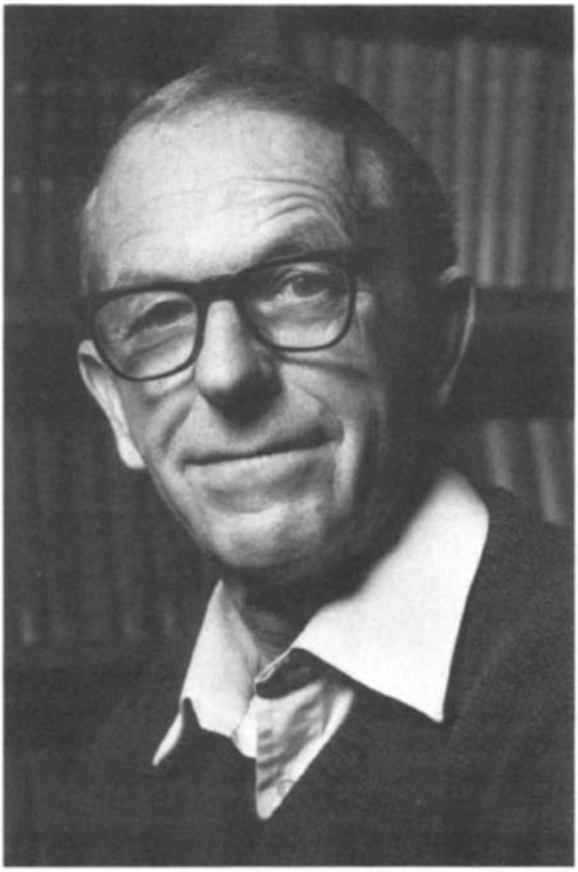

Ann. Rev. Biochem. 1988. 57:1–28

SEQUENCES, SEQUENCES, AND SEQUENCES

Frederick Sanger

Retired from Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge. CB2 2QH, England

CONTENTS

INTRODUCTION

These prefatory chapters are usually accounts of biochemists' experiences in research, teaching, and administration. In my case the last two are easily dealt with as I have done hardly any and have indeed actively tried to avoid both teaching and administrative work. This was partly because I thought I would be no good at them, but also out of selfishness. I do not enjoy them, whereas I find research most enjoyable and rewarding.

Of the three main activities involved in scientific research, thinking, talking, and doing, I much prefer the last and am probably best at it. I am all right at the thinking, but not much good at the talking. "Doing" for a scientist implies doing experiments, and I managed to work in the laboratory as my main occupation from 1940, when I started as a PhD student with Albert Neuberger in Cambridge, until I retired in 1983. Unlike most of my scientific colleagues, I was not academically brilliant. I never won scholarships and would probably not have been able to attend Cambridge University if my parents had not been fairly rich; however, when it came to research where

0066-4154/88/0701-0001$02.00

experiments were of paramount importance and fairly narrow specialization was helpful, I managed to hold my own even with the most academically outstanding.

This article is a personal account of my research career, which has been concerned primarily with the development of sequencing methods. Most of the significant work has been summarized in a number of reviews and articles. In these there was, of necessity, a good deal of simplification and omission of detail, both for reasons of space and, sometimes, to make a good and logical story. With the passage of time even I find myself accepting such simplified accounts, with no detail, no failures recorded, and very little reference to the colleagues who have helped me. This article is perhaps an appropriate place to clarify things somewhat and to describe certain lesser known aspects of the work, some of which have never appeared in print.

When I arrived at the University of Cambridge as an undergraduate, I had to decide which three scientific subjects I should take. I had chosen two-and-a-half subjects and was looking through the list of "half" subjects when I noticed one I had not heard of before: "Biochemistry, supervisor Ernest Baldwin." The idea that biology could be explained in terms of chemistry seemed an exciting one, and this was amply confirmed when I met the enthusiastic Dr. Baldwin. My first real academic success was when I stayed on at Cambridge for an extra year (1939) to take the advanced course in Biochemistry. I left Cambridge after taking the final examination with no high hopes of having achieved a good result, so I was very surprised to read in the newspaper two weeks later that I had earned a first class degree. This raised the possibility of my becoming a PhD student and I wrote to the Biochemistry Department at Cambridge enquiring if there was any prospect of this. As I received no reply I decided to visit Cambridge to investigate and found that plenty of research work was going on in spite of the war and that there were several people who would be glad to have a PhD student (especially one who, like me, did not need any money). Furthermore, I was a conscientious objector and had been given exemption from military service. Much of the work in the department centered on enzymes and intermediary metabolism, and the main tools were Warburg manometers and methylene blue tubes. During the advanced course I had been supervised by E. Friedmann, for whom I had developed a great respect. He was a refugee from Germany and a chemist of the old school, and in biochemical studies he believed in isolation and structure and did not have much use for what he called "making nonsense with methylene blue tubes." I think it was partly his influence that made me choose to work with N. W. (Bill) Pirie, who was the protein expert in the lab. Bill's chief interest at that time was in making edible protein from grass. His method of dealing with PhD students was the fairly effective one of "throwing them in at the deep end" and letting them find out things for themselves. He

presented me with a large bucket of frozen grass extract and suggested that I investigate it. Unfortunately by the time it had thawed Bill had left the lab for a new job, but not before I had assimilated some biochemical wisdom from him and enjoyed his caustic comments about the shortcomings of other members of the lab. I then needed another supervisor and was fortunate in that Albert Neuberger, who was a postdoctoral fellow in the laboratory, was prepared to take me on. I regard Albert as my main teacher. The most important thing he taught me, both by instruction and by example, was how to do research. I shall always be grateful for his kindness and patience. He also had an extremely wide knowledge of biochemistry, which I admired and used but could never emulate. The subject of my thesis was the "Metabolism of lysine": nothing very profound, but through it I gained much experience in amino acid chemistry, a good introduction to "sequences."

The Department of Biochemistry was an interesting and exciting place in which to work. It had been founded largely for F. G. Hopkins, and most of the staff members were his disciples, but also leaders in their own fields. Biochemistry was then a new subject, and there was plenty going on and a lot of enthusiasm. It seems to me that the quality of research in a laboratory is very much dependent on the atmosphere of the place. I remember being impressed with the friendliness and enthusiasm of everyone with whom I came into contact. The researchers seemed to be interested not only in what they themselves were doing, but also in what I was doing, and they were forever talking about biochemistry with no time wasted on "small talk." I liked this sort of environment and found it stimulating, and I have been fortunate in that both the laboratories in which I have worked have had this kind of atmosphere. The Laboratory of Molecular Biology, to which I moved in 1962, we built up ourselves and it could scarcely have failed to generate excitement and enthusiasm, but that initial impetus seemed to survive at least until I retired, and there were none of the major personal frictions that can have such an adverse effect on the research output of a laboratory.

INSULIN

When I had completed my PhD, in 1943, and Neuberger had left the lab, I was given the opportunity to work with Professor A. C. Chibnall and his group supported by an MRC (Medical Research Council) grant. Chibnall had just been appointed Professor of Biochemistry in succession to Hopkins and was about to move to Cambridge from Imperial College in London with a number of collaborators. Their main research interest at that time was the amino acid analysis of proteins, especially of insulin. Insulin had been chosen both for its medical importance and because it was one of the few proteins that could be bought in a pure form. It was unlikely to have been chosen because

of its small size (M_r 6000), as this was not known then, but it was of course very fortunate for our later work that it was so small. At that time amino acid analysis was a painstaking undertaking, and in many cases it was difficult to get accurate results. The analyses obtained by Chibnall's group were probably the best published at that time.

My work on insulin was done about 40 years ago and, because it is associated with the beginning of protein sequencing, has already been described, with varying degrees of accuracy, in a number of reviews and books. Some (e.g. Ref. 1), give a rather romantic impression of me, all alone, resolving to determine the structure of insulin against the advice of my colleagues who believed that proteins were some ill-defined amorphous mixture that could not be studied by chemical methods. This impression, however, is entirely wrong. Although there had previously been papers suggesting this pessimistic view of proteins, it was not one held by those with whom I associated. I think we all followed the lead of Emil Fischer and his school, which depended entirely on a chemical approach to proteins. If anything, the errors were in the opposite direction, ascribing more stoichiometry to proteins than was actually the case. In Britain the chief proponent of stoichiometry was the X-ray crystallographer W. T. Astbury. On the basis of a few spots on an X-ray film he proposed detailed structures for the folding of protein chains, and these were for a while widely accepted. The theory of Bergmann & Niemann (la) was particularly popular. They suggested that residues of a particular amino acid were arranged at regular intervals along polypeptide chains and that the content of each amino acid could be expressed by the formula $2^m \times 3^n$. A useful outcome of this theory was the stimulation of interest in amino acid analysis. People would express their results in this form, and, because of the lack of accuracy of the methods and as most small numbers fit the formula, this was not too difficult. Chibnall's was probably the first group to have results accurate enough to cause the theory to be questioned (2).

So the environment in which I worked was entirely favorable, and my initial experiments were logical continuations of the group's work, and in fact were suggested to me by Chibnall. Had it not been for him I would probably have continued working on lysine metabolism. Chibnall et al (2) had found that the number of free amino groups in insulin was considerably greater than could be accounted for by the lysine content. They suggested that the excess was due to free α-amino groups, indicating that the chains were relatively short and therefore particularly suitable for chemical studies. Chibnall suggested that I try to obtain a quantitative estimation and identification of the amino acids on which these free α-amino groups were located, and that is how my work on insulin started. In fact, the work on insulin could be considered to have started somewhat earlier than that as Neuberger had done some pre-

liminary experiments on the end groups, and probably his interest and experience were passed on to me. There was no question of my setting out to determine the complete sequence of insulin, then or for a considerable time thereafter. Probably the only time that I felt any real pressure to do so was in the final stages when only the disulfide bridges and one amide group remained to be located.

According to reviews that I and others have written, I then developed the use of fluorodinitrobenzene (FDNB) as a general reagent for determining end groups in proteins (3). Although this is so, on looking through my old notebooks I find it was by no means that simple. Several methods had already been suggested for labeling free amino groups, although in only one case was an N-terminal residue identified. In 1935 Jensen & Evans (4) had detected phenylalanine at the N-terminus of insulin, using phenylisocyanate, which is a somewhat less reactive analogue of the Edman reagent (phenylisothiocyanate) (5). This phenylalanine residue was, I think, the first amino acid to be located in any protein. For a general method what we needed was a reagent that would react under mild conditions with amino groups to form a derivative that was stable to acid hydrolysis. It was also important to have a relatively simple method for the separation and identification of the substituted amino acids. For this purpose we were anxious to use the newly discovered partition chromatography (6). It was a much better fractionation procedure than any that had been used previously. It worked well with the acetyl amino acids (6) and was in fact being used in the group by G. R. Tristram.

Besides the isocyanates the main possibilities were sulfonyl chlorides (7) or substituted benzenes (8). Benzene sulfonyl chloride had been used and was readily available, but the products did not seem to have very suitable solubility characteristics. The first reagent I tried was methane sulfonyl chloride. The main reason for this choice was that I expected the methane sulfonyl amino acids to behave similarly to the acetyl derivatives on partition chromatograms. This they did, but I was unable to get any clear results and we next turned to the substituted benzenes. 2,4 dinitrochlorobenzene had been studied by Abderhalden & Stix in 1923 (8); however it only reacted at high temperatures, where there was some hydrolysis of peptide bonds. I used it to make some dinitrophenyl (DNP) amino acids, and found they could be fractionated well on silica gel partition chromatograms. One great advantage was that the DNP derivatives were colored yellow. At that time chromatography really was "chroma"tography: there were no really reliable fraction collectors so that it was important to be able to see the bands on the columns. Clearly the chloro compound was not reactive enough, but the corresponding fluoro or nitro compounds could be expected to be more reactive. I made a small amount of the 1,2,4 trinitrobenzene and Chibnall discovered that Dr. B. C. Saunders, who worked in the nearby Chemical Laboratories, had a supply

of the fluoro compound, and he kindly let us have some. Both reacted satisfactorily in the cold. We had a good supply of the FDNB and so it was used in further work. The first experiments were, however, somewhat discouraging. I tried out the reaction of FDNB with glycine, but when I applied the product to a column, instead of the expected single band of DNP-glycine, I obtained two bands in significant amounts. It seemed that the second product was diDNP-glycine, with two DNP groups on the one N atom. I had considerable difficulties getting over this problem, but it turned out that glycine was an exception: other amino acids did not give so much of the second product and a change to bicarbonate from phosphate buffer almost completely eliminated the effect.

When I applied the method to insulin, DNP-phenylalanine and DNP-glycine were identified (3). There was one residue of each per M_r 6000 (although at the time we thought the M_r was 12,000). This suggested that there were two types of chain. These were joined by disulfide cross links of the cystine residues, which could be broken by performic acid oxidation (9), and it was possible to isolate two fractions that corresponded to the two chains (10).

At this time (1947) I had an opportunity to visit the laboratory of A. Tiselius in Uppsala, Sweden. Tiselius was one of the best-known names in proteins as he had invented the analytical electrophoresis method that bore his name and was very widely used. I had shown that the cross-linking disulfide bonds could be broken by oxidation and was trying to find a way to fractionate the chains. There were really no good methods for products of that size, but Tiselius's colleagues were working on a chromatographic system based on adsorption on charcoal which looked rather hopeful. It was suggested that I spend some time working on this in Uppsala. It seemed an exciting opportunity to me, especially as England was still in the postwar period of austerity and Sweden was, comparatively, a land of plenty.

The experiments I carried out with the charcoal columns were not very satisfactory. Detection of the bands coming out of the bottom of the columns was by a rather sophisticated method depending on refractive index: one had to look through an eyepiece observing "fringes" whose positions depended on refractive index. The trouble was the fringes would disappear when the refractive index changed and only reappear in different positions when the refactive index settled down again. One thus got rather few points on the plot. The custom in the lab was to have skilled technical assistants do the practical work, and there was one technician who was considered particularly good at reading these columns, so I got her to do some readings for me. A frontal analysis method was used, so that each product should show up as a step on the plot. The technician did several experiments, and one time the plot had four sharp steps. I was quite excited by this because at the time we thought the

molecular weight of insulin was 12,000 and that there were four chains. I showed this to Tiselius, and he immediately suggested we send a letter to *Nature* under both his and my name and include all the preliminary work on oxidation of the chains that I had already done in Cambridge. I was rather shocked as he had not really contributed anything and I had never seen him working in the lab. I managed to persuade him that I was not quite ready to publish the oxidation work, but being his guest and very much his junior I did not protest too hard about publishing the one promising result we had. The resultant paper (11) is fortunately one of my lesser known ones, and it is the only one of which I am ashamed. There are of course only two chains in insulin. I think the four sharp steps were due to the expert technician's way of plotting her readings; if the fringes disappeared she would assume there was no change and only record a higher reading when they reappeared, thus producing artificial steps. Tiselius was really a very kind and charming person, and I am certainly grateful for all the support he gave me both then and later, but this experience did make me realize how lucky I was to have worked with Chibnall, who had allowed me to publish my paper on the amino groups of insulin by myself even though he had actually initiated the work and could justifiably have put his name on it. One bonus from my visit to Sweden was meeting R. L. M. (Dick) Synge, who was working there and who introduced me to zone ionophoresis on starch, which was the forerunner of other more successful ionophoreses such as paper or acrylamide gel.

Complete acid hydrolysis of DNP insulin yielded DNP amino acids, but if the conditions of hydrolysis were reduced partial hydrolysis resulted and DNP peptides were obtained. By studying the DNP peptides from insulin I was able to identify short sequences of four to five amino acids near the N-terminus (12). The fact that only two sequences were obtained made it fairly certain that there were only two types of chains, and, what was probably more important, showed that specific sequences could be determined in proteins by partial hydrolysis, thus eliminating any doubts that they were pure stoichiometric compounds that could be studied by the methods of organic chemistry. In most subsequent work, both on proteins and nucleic acids, partial hydrolysis has been the main general approach to all sequencing, and it is only relatively recently in DNA sequencing that it has been largely superseded. Although my first paper on the N-terminal residues (3) has been more widely quoted (13), I was myself more excited about this peptide paper (12), and indeed in retrospect the results seem to me to be of greater significance. It was the first time that sequences were located in a protein molecule, and we could draw important conclusions about stoichiometry. Also it was done on my own initiative, whereas the first paper relied more on Chibnall's stimulus.

Apart from my two teachers, Neuberger and Chibnall, my first actual collaborator was Rodney Porter. Although a little older than I was, he was

officially my PhD student. This was because he had been in the army during the war, while I had been working in the lab. Rodney was not the sort of person one could supervise, but he was very good as a collaborator and I think I learned more from him than he did from me. Probably the main thing he taught me was a certain lighthearted attitude toward research, which makes it more enjoyable and, provided standards of complete integrity and honesty are maintained, can do no harm. He joined me just after the development of the DNP method, when we were extending it to peptides and to other proteins. Rodney had just developed an interest in γ-globulin and wanted to study its end groups. I warned him that it was a complex mixture and that it was not worth trying to determine its end groups. However, this did not stop him, and to everybody's surprise he got a sensible, simple result; and from then on he was inseparable from γ-globulin, to the great benefit of science.

During this period we were consigned to a lab in the basement next to the experimental rats. In spite of these rather unsavory surroundings, it was probably the most enjoyable lab I have worked in. We shared it with Kenneth Bailey and his PhD student, S. V. (Sam) Perry, who had also been in the army and had been at the University of Liverpool with Porter. Thus Bailey and myself, two reserved and serious-minded scientists, were thrown together with these two wild boys fresh from the army. It certainly kept us awake. Both had vocal tendencies: Perry's favorite tune was "Mighty mountain" rendered after the manner of Paul Robeson, and Porter's was "Nobody knows the trouble I've seen." It was a good atmosphere and at the same time we accomplished quite a lot. The only serious black spot was when a flask exploded in front of Rodney's eye and damaged the cornea. Although the sight was not completely lost, I think he never had good vision in that eye.

Although the DNP method was an improvement on previous ones and important results were obtained using it, there were certain disadvantages, the most serious being that the DNP amino acids were not completely stable in acid and a correction factor had to be used. This factor differed with the various amino acids and for some (e.g. glycine) was fairly large; in fact in the case of hemoglobin this problem led to an error in the estimation of the free amino groups. I was therefore still interested in trying to develop a better method. The sulfonamide (SO_2NH) bond is considerably more stable than the DNP, so that it seemed a sulfonyl chloride might be an improvement, and I was anxious to try a more highly colored material to increase the sensitivity. The reagent on which I spent quite a lot of time was dimethylaminophenyl-azo-benzene sulfonyl chloride (or helianthyl chloride). The helianthyl amino acids gave beautiful red bands on the chromatograms and appeared to be stable to acid, but we did not have much success when we tried this reagent with proteins. There were probably some undesirable side reactions with the azo group. There were also complaints from other occupants of the lab when

their biological preparations all became bright red. Eventually we gave it up, but more success was obtained later by Hartley & Gray (14) with another sulfonyl chloride, diamino naphthalene sulfonyl chloride (dansyl chloride). The dansyl amino acids were satisfactorily stable and their strong fluorescence made them easy to detect and assay at low concentrations.

The DNP method was fairly widely used after its development, but is now completely superseded by other methods, especially the Edman procedure (5), which has the advantage that it can be used not only for determining N-terminal residues but also for extensive sequential degradations of polypeptide chains. It is interesting that the Edman method was developed about the same time as the DNP method but at first was not used much. The main reason for this was probably the color of the DNP derivatives, which made them much easier to handle at a time when reliable fraction collectors and micro methods of assay were not available.

For sequencing purposes the DNP method was very limited as it could only be used for a few residues at the N-terminus, and the more general approach of subjecting a whole chain to partial hydrolysis seemed necessary to extend the sequence. The main technical problem was fractionation of the complex mixture of peptides produced. Indeed in almost all sequencing work fractionation has been a crucial factor, and progress in sequencing has often depended on progress in fractionation methods. With the insulin chains we were again lucky in that the method of paper chromatography had just been developed by Martin and his colleagues (15), and had been used to determine the pentapeptide sequence of gramicidin S (16). The fractionation of small peptides was far superior to anything that had been achieved previously, and it seemed something of a miracle at the time to see these hitherto intractable products separated from one another on a simple sheet of filter paper. Using this method we were able to isolate small peptides from acid hydrolysates, some of which could be fitted together to deduce larger sequences (17, 18). However, with acid hydrolysates alone we could not obtain the complete sequence of either chain, and what we clearly needed were longer peptides. A more specific method of hydrolysis was required and the obvious choice was proteolytic enzymes. Hitherto we had been reluctant to use enzymes because it was rather generally thought that they might synthesize bonds as well as hydrolyze them, and thus sequences could be rearranged. Enzyme reactions were known to be reversible, and a good deal of work was being done on the synthetic activity of proteolytic enzymes, since it was considered that this might be involved in protein biosynthesis. In spite of these warnings, we decided to risk it and to try using proteolytic enzymes, and we were able to obtain the complete sequence of both chains (19, 20). Knowing that we already had much of the sequence from the acid hydrolysis results and that there were no anomalies between the two methods convinced us that there had

been no rearrangement. In all subsequent work proteolytic hydrolysis has been the method of choice, and in fact rearrangements have been shown to occur during acid hydrolysis (21).

The work on the sequencing of the two chains owes much to my two collaborators—Tuppy and Thompson. Hans Tuppy was from Vienna and was one of the hardest workers I have known. Just after the war there were only limited facilities for research in Austria, and he was delighted to come to a well-equipped lab where there was no limitation on the amount of work he could do. We were at the time collecting peptides from acid hydrolysates of the chains of insulin. He took the longer phenylalanyl chain, and I worked, among other things, on the glycyl chain. Much of the work required the use of paper chromatography; the chromatography room was at the other end of the basement corridor, and a common sight at that time was Tuppy walking at full speed along that corridor bearing chromatograms. He never ran, but to keep up with him anyone else had to run. He only stayed a year, but by the end of that year the phenylalanyl chain was practically finished and the glycyl chain had to wait for my next collaborator, E. O. P. (Ted) Thompson from Australia. I have worked with several Australians and have always found them easy to collaborate with. They seem to have a certain practical "no nonsense" view of research that takes the attitude, "If there's an experiment to be done, get on and do it." Ted was no exception and we had great fun working together. Among other things we finished up the glycyl chain, which I think really was harder than the longer phenylalanyl chain, and his persistence in tracking down the final amide group was ultimately rewarded.

Once we had determined the complete sequence of the two chains, the only remaining problem was the arrangement of the disulfide bridges. This proved particularly difficult and probably involved as many man-hours, and more frustrations, than the rest of the work put together. The idea was to hydrolyze the intact insulin, isolate cystine peptides, oxidize them, and see which cysteic acid peptides were produced. The results were, however, very equivocal, and it eventually transpired that there was an interchange reaction taking place at the disulfide bonds during acid hydrolysis (21a). One bond between the two chains could be identified using enzymic digestion, but the other two occurred in a Cys-Cys sequence in the glycyl chain, and no enzyme would split the two half-cystine residues. Clearly, in this respect insulin had been a bad protein to choose; hardly any other protein would have given us this problem. It was quite a long time before we found conditions for avoiding this interchange reaction during acid hydrolysis and were able to complete the structure (22).

Insulin was the first protein to be sequenced (22, 23), but its chains were relatively short, and it is doubtful if the methods used would have succeeded with larger, more normal, chains. Further fractionation methods were needed

and the most important was that introduced by Moore & Stein (24, 25): ion-exchange chromatograms to separate the large peptides from proteolytic digests. They also used ion-exchange columns to develop an accurate method of amino acid analysis, and this, though considerably modified and improved in detail, is still the method of choice today. Using these techniques they were able to determine the complete sequence of ribonuclease (120 amino acids) (25), and subsequent work on protein sequence has more closely followed their techniques than the less quantitative methods employed for insulin.

THE LEAN YEARS

After completion of the insulin sequence around 1955, there followed a number of "lean years" when there were no major successes. I think these periods occur in most people's research careers and can be depressing and sometimes lead to disillusion. I have found that the best antidote is to keep looking ahead. When an experiment is a complete failure it is best not to spend too much time worrying about it but rather to get on with planning and becoming involved in the next one. This is always exciting and you soon forget your troubles. One disadvantage of this attitude, though, is that it makes the writing of an article of this sort difficult, because if you are always thinking about the future you tend to forget the past.

Toward the end of the insulin work people would ask me what I would do after I finished insulin. My usual reply was that knowing the structure was only a beginning and that the next thing was to find out what it meant and how the insulin worked. I have always been rather unsuccessful in predicting the future, especially in my own research, and although I did try to do some experiments toward this end, my interest began to go in other directions, and perhaps the sequencing bug had already taken hold of me.

During these "lean years" there were a number of events that are worth recording as they had an important effect on the subsequent work on nucleic acids, although they may not have led to publishable results. The first was initiated by the visit of Chris Anfinsen on sabbatical leave to our laboratory in 1954. Chris had been using, and was very enthusiastic about, the new technique of labeling with radioactive isotopes, and in discussion with him I caught some of his enthusiasm. Previously I had assumed that isotopes were in the realm of the physicists and that the apparatus and techniques would be beyond my means. But I learned that this was not the case and that already a number of radioactive substrates were available. Together we labeled a rat with ^{35}S in an attempt to prepare radioactive insulin. The experiment was not a success, but nevertheless I became a complete convert to the use of radioactive isotopes and to the autoradiographic technique, which was a very powerful and simple tool especially when used in conjunction with paper

fractionation methods (paper chromatography and paper ionophoresis). After the introduction of ion-exchange chromatography for fractionating peptides (24), paper methods became less fashionable and were largely replaced by columns and fraction collectors, which were more accurate and reproducible. I was, however, never really converted to columns and retained a certain affection for the paper methods. It seemed to me much more exciting to develop a two-dimensional paper separation and to see all the spots spread out before one than to collect innumerable fractions from a column, analyze them all, and end up with what was usually the expected clear, clean data. There was more, if less clean, data on the paper, and hence more opportunity for unexpected and exciting findings. Although my preference for paper methods was to some extent sentimental, I think it was scientifically justified if only because of its subsequent successful application to RNA.

Somewhat more successful than the ^{35}S-labeling were later experiments with ^{32}P. Some proteins, notably ovalbumin (26), could be labeled with [^{32}P]-phosphate, and some enzymes, such as trypsin and chymotrypsin (27, 28), could be labeled in their active centers with [^{32}P]-diisopropyl phosphofluoridate (DFP). In this way just one or two residues in a whole protein were labeled, and if a partial hydrolysate were fractionated on a two-dimensional paper system and autoradiographed a relatively simple picture was obtained. I was particularly interested in developing methods for deducing the sequence of peptides without resorting to the standard procedure of carrying out complete amino acid analyses, which were somewhat time-consuming and tedious. The ^{32}P-labeled proteins were particularly suitable for trying this type of approach. For instance, information could be obtained by comparing autoradiographs of ionophoreses of partial acid hydrolysates of ^{32}P-labeled proteins. In this way it was shown that the sequence around the reactive serine in elastase was the same as that in trypsin and chymotrypsin (27). Each labeled residue gave a specific pattern of bands or "fingerprint," and techniques were worked out for identifying the sequences by autoradiographic methods without carrying out an amino acid analysis. The main rationale behind this approach was that the radioactive peptides were usually heavily contaminated with nonradioactive substances, but they were radiochemically pure and thus the autoradiographs gave a simpler and more interpretable fingerprint. The relationships of the different bands in a fingerprint could be determined by subjecting them to a further partial hydrolysis. Information about charged amino acids was obtained from the ionophoretic behavior of the peptide bands, and about neutral amino acids from chromatographic behavior. Treatment with phenylisothiocyanate or with periodate gave information about the free amino groups. Using this type of method it was often possible to determine a sequence of four or five amino acids around a specific labeled residue (27–31). Particularly rewarding at this time was a conversation I had

with Cesar Milstein, who was working at the other end of the corridor on phosphoglucomutase. This led to a collaboration in which we labeled its active center with ^{32}P and determined a sequence of five amino acids (30). Although this was our only collaboration, my continued association with him has meant a great deal to me.

Even more ambitious but less successful was a means I devised for determining the complete sequence of a protein by autoradiographic methods. The idea was to prepare 20 samples of the protein, each labeled with a different amino acid. Ovalbumin was chosen to study as it was relatively easy to label in isolated oviduct tissue. The samples were then digested with a proteolytic enzyme and the digests fractionated in parallel by paper electrophoresis and an autoradiograph prepared. Assuming that a peptide was pure, its amino acid composition could be deduced from the darkness (radioactivity) of its band in the separate channels. The whole paper was then to be subjected to the Edman procedure and washed to remove the phenylthiohydrantoin, and another autoradiograph was to be prepared. The N-terminal residue of each peptide should then be identified as the band missing or weaker in the second autoradiograph. By doing further Edman degradation on the paper, further sequences of the individual peptides were to be determined. The idea was certainly rather wild, and it did not show much sign of working. However, if it had worked it would have been exciting and would have been an alternative to the standard partial hydrolysis approach, and it might have been quicker and simpler to carry out. We spent quite a lot of time working on this idea and the question arises whether this was really justified. At that time I was in the fortunate position of having a permanent research appointment with the (British) Medical Research Council, and was not under the usual obligation of having to produce a regular output of publishable material, with the result that I could afford to attack problems that were more "way out" and longer term: in fact, as few others could adopt this approach, I felt under some obligation to do so. This type of work also suited my temperament. I like the idea of doing something that nobody else is doing rather than racing to be the first to complete a project, and I prefer to concentrate more on the practical work itself than on its ultimate outcome. Although this work was unsuccessful, the proposed method bears some resemblance to the DNA sequencing methods we finally developed, and so probably our efforts with proteins were worthwhile, and they give some indication of how our ideas about sequencing were developing. My only regret about this work is that it seemed rather unfair to my excellent assistant, Biddy Segall, who devoted so much time and skill to this project but whose name never appeared on a paper.

During most of the work on insulin I was supported by a Beit Fellowship; this lasted seven years (a Junior Beit, a fourth year, and a Senior Beit), and I

was certainly very grateful for this continued support, which allowed me to devote all my time to research in the Biochemistry Laboratory. Thereafter I was supported by the Medical Research Council—initially as an external member of the scientific staff. In 1962 I moved to the new MRC Laboratory of Molecular Biology where I joined forces with the group led by Max Perutz from the Cavendish (Physics) Laboratory; this included, among others, Francis Crick, Sydney Brenner, and John Smith. The new laboratory was divided into three more or less self-governing divisions, with Max as Chairman. One division (the Division of Protein Chemistry, which later became the Division of Protein and Nucleic Acid Chemistry) was allotted to my group. There was much more space than I had ever had previously, which meant that I could expand considerably. As I wanted to continue doing experimental work, I had no ambition to have a large team working under my direction, but this gave me a chance to bring together a number of people with similar interests in protein chemistry to make a viable group—a few collaborating directly with me and others either working independently or with a more senior member of the group. Initially the senior members were Ieuan Harris and Brian Hartley, who had been fairly closely associated with me in the Biochemistry Department, and each had his own group. We were later joined by Cesar Milstein. The laboratory rapidly acquired a good reputation so that we were able to attract plenty of good postdoctorals and PhD students, which greatly helped in maintaining the enthusiasm and momentum of the lab. At first there was plenty of room and everyone had a large bay to himself, which seemed a great luxury; however this did not last long and soon the laboratory was as overcrowded as most. In fact the work seemed to go better in the overcrowded conditions.

I think the move to the new laboratory was probably an influence in my conversion to nucleic acid sequencing. Previously I had not had much interest in nucleic acids. I used to go to Gordon Conferences on Protein and Nucleic Acids when the two subjects were bracketed together, and would sit through the nucleic acid talks waiting to get back to proteins. However, with people like Francis Crick around, it was difficult to ignore nucleic acids or to fail to realize the importance of sequencing them. An even more seminal influence was John Smith, who was the nucleic acid expert in the new laboratory and who was extremely helpful to me, so that I could turn to him for advice in this new field. He might well have felt I was "muscling in" on his area, but he did not, and was always ready to give me the benefit of his long experience.

RNA

Initially the sequencing of nucleic acids seemed much more difficult than that of proteins, and little progress had been made until the early 1960s. This was

partly because of the lack of pure small substrates on which to develop methods, and partly because of their composition. For nucleic acids with only the four monomers, the interpretation of results would be expected to be more difficult than for proteins with their 20 amino acids, and larger degradation products would have to be isolated to give significant overlaps for sequence deduction. On the other hand, only having four components would make the final analyses much easier. To begin with the difficulties of interpretation were predominant, but as the techniques have improved and larger molecules have been studied the question of analysis has become more important, so that today nucleic acid sequencing is much quicker and simpler than protein sequencing. Whereas for proteins much of the time is spent in identifying and estimating the amino acid end products, it has now been possible to evolve methods for DNA in which this final analytical step is completely eliminated.

The first nucleic acid to be sequenced was in fact the first small RNA to be purified. This was the alanine transfer RNA, which was purified by Holley and his collaborators (32). They used sequencing methods that were very similar to those in use with proteins: partial hydrolysis with enzymes, and fractionation of the products on ion exchange columns. They introduced the use of the enzyme ribonuclease T_1, which, because of its unique specificity for splitting after G[1] residues, has been used as the main degradative method in most subsequent work.

Our own early efforts in RNA sequencing were directed toward the development of more rapid and simpler fractionation techniques and were influenced by our recent experience with ^{32}P-labeled proteins. Since every nucleotide contains a phosphorous atom, RNA is particularly suitable for ^{32}P-labeling in vivo, making very sensitive detection by autoradiography possible. The use of columns for fractionation was somewhat laborious, and it seemed that the simplicity and resolving power of two-dimensional techniques would offer considerable advantages for separating the complex mixtures of small nucleotides that one would obtain from partial degradation of larger RNAs. In general, oligonucleotides did not separate well by simple paper techniques (partition chromatography or ionophoresis), but we were able to develop a two-dimensional method that used ionophoresis on cellulose acetate followed by ionophoresis on ion exchange paper, and this proved considerably better than any of the previous methods (33).

I am sometimes asked such questions as "What was the most exciting moment of your scientific career?" or "What did it feel like when you discovered. . . .?" Such questions are usually prompted by the popular idea

[1]Where there is no possible ambiguity in the text the terms A, G, C are used for both the ribo and deoxyribonucleotides; otherwise rA, rG, rC, U are used for the ribo and dA, dG, dC, T for the deoxyribonucleotides.

that scientific progress depends on sudden breakthroughs or moments of inspiration (cf Archimedes jumping out of his bath, or Watson and Crick discovering the double helix). For me, however, progress does not seem to go like that, and I find it hard to remember moments of sudden exhilaration. Although there were such moments, they were probably more often associated with small and gradual advances; and these were more numerous and therefore gave more enjoyment than would have been the case if there had been sudden big leaps forward. However, I do have a clear recollection of one occasion when Bart Barrell, who usually developed the day's autoradiographs first thing in the morning, came into my lab brandishing a beautiful sheet of film with clear, round, well-separated spots on it. This was certainly exciting after the streaky, unresolved pictures we had been getting before.

Figure 1 *(left)* shows a diagram of an autoradiograph of a ribonuclease T_1 digest on ribosomal RNA. In such a digest all products have G at their 3' ends; there are three dinucleotides and nine trinucleotides, and most of these occupy unique positions so that in subsequent experiments 10 oligonucleotides could be identified simply by their position. This avoided a good deal of final analysis, and was a first step toward our ambition of eliminating the analytical step in nucleic acid sequencing. Furthermore, the composition of larger nucleotides could be determined from the "graticules," shown in the righthand side drawing. All fragments contain one G. The residue that has the greatest effect on the mobility is U, so that the pattern is divided into separate, distinct groups of spots; all oligonucleotides with no U are in the fastest

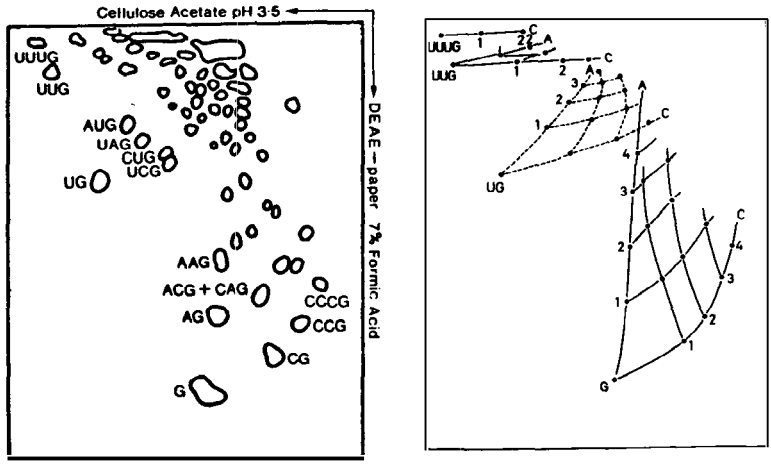

Figure 1 *Left:* Drawing of autoradiograph of two-dimensional separation of ribonuclease T_1 digest of ribosomal RNA. *Right:* Graticules showing the relation of the composition of an oligonucleotide to its position on the autoradiograph.

moving group, those with one U in the next, and so on. Within the groups it is possible to draw a graph or graticule in which one axis gives the number of C residues and the other the number of A residues.

Following this general approach we developed a number of methods for studying the isolated nucleotides. This was in a way an extension of the work on ^{32}P-labeled peptides, though much easier for RNA with only the four components to be considered. One such method that was used with oligonucleotides from pancreatic or T_1 ribonuclease digests was to subject them to partial digestion with a 5' exonuclease and to run the products on ionophoresis on DEAE-cellulose paper at pH 1.9. Sequential degradation from the 5' end will give a mixture of fragments all with the same 3' end but differing in their 5' ends. On the ionophoresis these will be arranged in size order. From the relative positions of two adjacent bands it was possible to identify the nature of the nucleotide by which they differed. Figure 2 shows an example of this approach with the oligonucleotide GAAGC from a pancreatic ribonuclease digest. We defined a value M as the distance between two adjacent oligonucleotides, x and y, divided by the distance of x (the larger) from the origin. If the two differ by an A the M value is 0.5–1.0, whereas in the case of G it is 1.2–3.1. In Figure 2 *a* is the unchanged material and *b* the

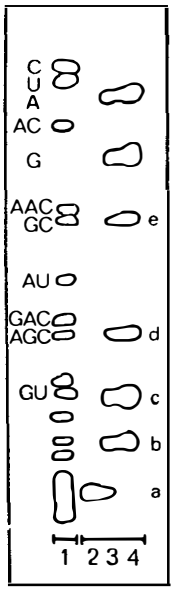

Figure 2 Sequence determination by partial digestion with 5' exonuclease of the oligonucleotide GAAAGC. 1. A pancreatic digest of ribosomal RNA used as marker. 2. Untreated GAAAGC. 3. and 4. GAAAGC digested with 5' exonuclease for 30 and 60 min, respectively.

first degradation product. The M value is 1.9 and therefore the first residue must be G. The M value for *b* and *c* is 0.55, so the second residue is A, and so is the third. Bands *d* and *e* are recognized as AGC and GC by their unique position relative to known markers.

Another useful method was the "wandering spot" technique, which was also sometimes known as "walking the graticule." This was essentially a two-dimensional extension of the above technique and was used mainly with DNA. It was first worked out by Victor Ling to sequence large depurination products (34). A two-dimensional system was devised in which the fragments from an exonuclease digest were arranged according to size, so that each spot differed from the one next to it by a single nucleotide. The system was also arranged to that the relative positions of two neighboring spots depended on the nucleotide by which they differed, as shown in the inset in Figure 3. The method was extended for use with more complex digests, but it was not possible to distinguish A and G with absolute certainty. Figure 3 shows the application of the method to a dodecamer that contained only C, T, and A. The diagram on the left shows a 3' exonuclease digest from which the sequence TTACCATT can be read, and that on the right a 5' exonuclease digest giving the sequence CTTATTAC. Combining the results gives the full sequence CTTATTACCATT.

A suitable substrate for trying out our new RNA sequencing strategies was the 5S ribosomal RNA of 120 residues, and this was completely sequenced mainly by G. G. Brownlee; at that time it was the largest RNA to have been

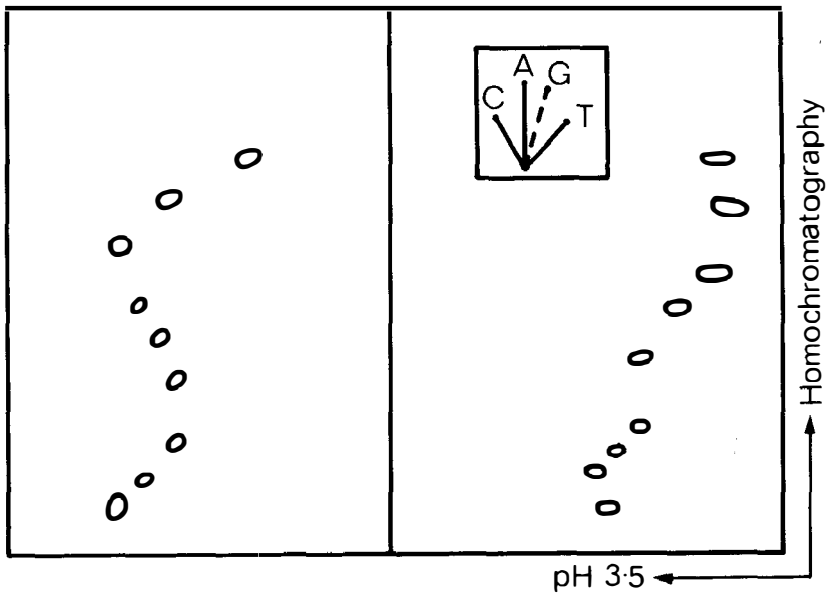

Figure 3 The "wandering spot" method applied to the oligonucleotide CTTATTACCATT.

sequenced (35). We were also able to determine sequences in the RNA bacteriophage R17 that were coding for a part of the head protein whose amino acid sequence was known (36, 37). When we started sequencing RNA one of the main objectives was to try to "break the genetic code." However this was done by entirely different in vitro translation techniques before the sequencing methods were sufficiently advanced. Nevertheless the results with R17 were a useful confirmation of its correctness.

During the work on nucleic acids I had more collaborators than previously and I have been extremely fortunate in the people who have worked with me. Many have been American postdoctorals. We used to get rather many applications, and it was always difficult to choose from such a glittering group. The main problem is that it is not only academic brilliance you are looking for. For a successful collaboration one of the most important things is to have someone you like and can get on well with, and references are often not helpful in this respect. When I was considering John Donelson, one of the referees had added a postscript saying "I think you will enjoy his sense of humour." This proved to be the case and I found the comment helpful. For successful collaborative work there must be complete understanding and trust, and a sharing of ideas and resources. If frictions arise then usually the work suffers. Usually we have had a good atmosphere in the group, but at one time when we were starting our work on DNA we had a relatively large number of people and I noticed that a certain amount of rivalry and bad feeling seemed to be building up. This worried me rather as I did not know how to deal with it, never having had the problem before, so I largely "acted ignorant" and took no notice of it. I think this was probably the best policy; by assuming that they were all friends they were shamed into recognizing the stupidity of their behavior.

The two people to whom I owe most during the nucleic acid work are my two personal assistants, B. G. (Bart) Barrell and Alan Coulson. At about the time I was changing from proteins to RNA I was looking for a new assistant and there were very few applications. The only serious possibility was B. G. Barrell, a young fellow fresh from school. He did not have the academic qualifications we were looking for, but he had plenty of enthusiasm. He came for an interview; I took to him and decided to risk it. He already knew a surprising amount about molecular biology, which he had clearly found out for himself—probably at the expense of his more orthodox school work. He came from Gloucestershire, which was also my own home county, and spoke with a nice thick Gloucestershire accent. He soon became a first-class sequencer and subsequently took charge of his own group. He gradually took over more and more of the nucleic acid sequencing work, especially the actual determinations, while I concentrated more on methods. This meant that I needed a new assistant, and I was fortunate to find Alan Coulson, who was my main collaborator in the later DNA work, and who in his own quiet

way contributed to every aspect of it. Hitherto I had preferred working with the more outgoing, extroverted types whose personalities were in marked contrast to my own. I suppose I admired them and found them stimulating. With Alan it was different; we were both quiet by nature. Perhaps as I became more mature I felt less need for external stimulation, though there were always plenty of noisy people around in the lab.

DNA

The first attack on DNA in our group was made in the mid-1960s by Kenneth Murray, who developed some two-dimensional systems for fractionating and analyzing small oligodeoxyribonucleotides (38). Small sequences could be obtained, but at that time there was no way of locating them in the chains. The main problem with DNA was the very large size: the smallest pure DNAs that were available were the genomes of the single-stranded bacteriophages (such as øX174, which will be referred to as øX) of about 5000 nucleotides, and these were rather large for testing out methods. Another difficulty was the absence of suitable degradative enzymes. The sequencing of RNA depended very much on the use of ribonuclease T_1 with its specificity for one residue (G). No enzyme with an analogous specificity exists for DNA. The restriction enzymes, which have since played a very important part in DNA sequencing, had not yet been developed. We did however, around 1973, carry out some experiments using similar techniques to those employed for RNA and were able to determine a few sequences of about 50 residues (39–42). However, the methods were both slow and laborious, and it seemed that if we were really going to be able to attack the vast sequences of genetic materials, then an entirely new approach was needed. As an alternative to partial hydrolysis, some work had been done on the possibility of using copying techniques for sequencing, particularly by C. Weissmann and his colleagues (43). The RNA bacteriophage Qβ contains an RNA polymerase that copies its own RNA, and they had developed elegant pulse-labeling techniques for the RNA and for deducing sequences. An obvious enzyme to choose for copying DNA was DNA polymerase, which had already been used by Wu & Kaiser in 1969 (44) to determine the sequence of the "sticky ends" of bacteriophage λ. This dodecanucleotide sequence was in fact the first bit of DNA to be sequenced (45).

An elegant approach to specific digestion of DNA, which could be combined with a copying procedure, was suggested by Berg, Fancher, & Chamberlin in 1963 (46). In normal conditions the only substrates for DNA polymerase are the deoxyribonucleoside triphosphates, but if magnesium is replaced by manganese in the medium, ribonucleoside triphosphates can be

used. Thus, for instance, if incubation is with one ribonucleoside triphosphate (say rCTP) and three deoxyribonucleoside triphosphates, a DNA chain is built up with all the dCs replaced by rCs. These bonds are labile to alkali or ribonuclease, and so it was possible to prepare a C-specific digest, and similar digests to split at other residues. The method looked very promising, and we did a good deal of work on it. DNA polymerase requires a single-stranded DNA as template and a primer that is another piece of single-stranded DNA that base-pairs with a specific region on the template. As the template we could use the single-stranded DNA of bacteriophage f1. Our intention was to use as primer an oligodeoxyribonucleotide having a sequence that could be predicted from the known amino acid sequence of the coat protein of the bacteriophage. The main question was how to make such a primer. At the time Khorana and his colleagues were developing methods for synthesizing oligodeoxyribonucleotides, but the methods were rather slow and specialized and we had no experience. Fortunately when attending a meeting at about that time I met Hans Kössel from the University of Freiburg, who had worked in Khorana's lab. I found that he had the same idea and was thinking of synthesizing the same primer, so we decided to collaborate. It took him and D. Fischer more than a year of hard work to make the octanucleotide primer. If we had waited a year or two restriction enzymes would have been available and we could have used a restriction enzyme fragment as primer. The ribo-substitution method worked reasonably well and we were able to determine a sequence of about 80 nucleotides (39, 40). Our plan was to make another primer with a sequence corresponding to the 3' end of the one we had determined and get another 80 residues, and so gradually work round the genome. This would obviously still be a slow process with many fragments to be analyzed, as well as primers to be synthesized. Fortunately, the research developed in a different direction.

In the above experiments we wanted to obtain very highly labeled DNA, and therefore used the radioactive substrate at high specific activity and in low concentrations. In such experiments (using ^{32}P-ATP) it was frequently observed that the DNA products formed ended at a position before an A would have been incorporated. Presumably the enzyme was running out of ATP. This suggested a new approach to DNA sequencing. If one could produce a mixture of chains all having the same 5' end (corresponding to the 5' end of the primer) and finishing at the 3' end at positions corresponding to the A residues, determination of the relative sizes of all these chains should give a measure for the relative positions of the A residues, and this, combined with similar data for the other three nucleotides, is all one needs for the complete sequence determination. Various methods were studied for fractionating the products of synthesis according to size, and one of these was found to be satisfactory. In fact it proved to be much more efficient than we

had ever imagined would be possible. This was electrophoresis on acrylamide gel. If carried out in suitable denaturing conditions, oligonucleotides up to over 300 residues long could be separated according to size, the smaller ones migrating more rapidly on the gel than the larger ones, with each product clearly separated from its neighbors, which differed in size by only one nucleotide.

This new approach to DNA sequencing was I think the best idea I have ever had, being original and ultimately successful, so I have attempted to describe its development in some detail, but on reading it through I must confess that I am by no means certain that it really did happen like that, I certainly do not remember having the idea, whereas I do remember doing some preliminary experiments and discussing it with Alan Coulson and John Donelson, I have a feeling that the above account may have originated to some extent from my attempts to explain the method in a simple way when giving lectures, and that subsequent frequent repetition resulted in its being established as part of my "official," but perhaps not actual, memory.

The method as first conceived, using low concentrations of the substrates, was not very satisfactory, but we did not wish to give up an approach that seemed so promising and, after introducing various modifications of the conditions, two analogous techniques (which we called the "plus and minus" method [47]) were developed and proved to be a much more rapid way of sequencing DNA than anything that had gone before. Most of the sequence of øX was determined using this method. However, a much more efficient and reliable method soon superseded this. For the approach to work we needed to be able to prepare a mixture in which the various end products (all with the same residue at the 3' end) were present in about equal amounts. In the plus and minus method this was not the case, and the sizes of the products were distributed over a rather narrow range so that only relatively short sequences could be determined from one incubation. Another way of achieving the same effect was suggested by the work of Kornberg and his colleagues (48) with analogues of the normal DNA polymerase substrates, which act as chain-terminating inhibitors. One of these was 2',3'-dideoxythymidine triphosphate (dideoxy TTP), identical to the normal TTP but lacking a 3' hydroxyl group. It is a substrate for DNA polymerase, though not such a good one as TTP. Once incorporated the chain no longer has a 3' hydroxyl group and so cannot be extended. If a reaction with DNA polymerase is carried out using the four normal nucleoside triphosphates and a suitable concentration of dideoxy TTP, one would expect on long incubation to end up with a mixture of products all with dideoxy T at their 3' ends. The main problem initially was that these analogues were not available. However, at a meeting in Germany I discovered that K. Geider had made some dideoxy TTP, and he very kindly let me have some. The first experiment we did with it gave a beautiful autoradiograph

with sharp bands of equal intensities extending over a long sequence. Clearly this would be much better than the plus and minus method. The problem then was to obtain the other three dideoxy NTPs, which had never been synthesized. I was told by a representative of one of the drug companies that his company was going to make them. After waiting about a year, I met him again and he casually told me that his company had decided not to, so Alan Coulson and I decided to make them ourselves. We had no experience, but we had some expert chemists in the lab to advise us, and quite enjoyed this different type of activity. Eventually we succeeded in preparing all four and the method worked well (49). It was used first to complete the øX sequence. In this case the single-stranded phage was used as template, and restriction enzyme fragments from the double-stranded (replicative) form of øX were used as primers. We naturally wanted to extend the method to other and larger DNAs; however there were two major problems. The first was the eternal one of fractionation, which has always played such an important part in sequencing studies. As the methods became more speedy the preparation of restriction fragments for primers became more and more of a limiting factor. This had been done by acrylamide gel electrophoresis. With øX reasonably pure products could be obtained from a single run, but with larger DNAs this was clearly going to be more of a problem. The other difficulty was that the method requires single-stranded DNA as a template. This was fine for øX whose normal form is single-stranded, but most DNAs are double-stranded, and it is difficult if not impossible to prepare the single strands. Various means were tried to overcome this problem, but the best was cloning the DNA in a single-stranded phage vector. Not only is this a means of obtaining single-stranded DNA, but it also solves the fractionation problem as fractionating is by cloning, which is really the ultimate method of purification and can be applied to a mixture of any complexity. The most efficient method was devised by Messing and his colleagues (50, 51) and is now the quickest method for sequencing DNA of almost any size. The vector used is the double-stranded form of the single-stranded bacteriophage M13, modified to contain a site into which fragments of the DNA to be sequenced can be inserted. The mixture of recombinants is then plated out and grown. This is essentially the fractionation method. Each plaque is derived from a single molecule and is therefore pure. The phage from the plaque is then grown and the single-stranded DNA isolated from it. This is used as the template for the dideoxy sequencing, using as a primer a synthetic oligonucleotide that is complementary to a part of the M13 vector that flanks the insert in such a way that the insert is copied by the DNA polymerase. In this way the same synthetic primer, which can be prepared pure in relatively large amounts, is utilized to prime the different recombinants.

The method is normally used as a random procedure. Originally restriction

enzyme digests were used for inserting into the vector, but this was subsequently found to be unnecessary and better results are now obtained by random cleavage, usually by sonication (52).

I suppose the dideoxy method can be regarded as the climax of my research career and the fulfilment of an ambition that had gradually been forming as I became more and more involved in sequencing. It is of course very exciting and gratifying to read of the method now being used in many laboratories and of vast regions of the genomes of both simple and higher organisms becoming exposed in the form of sequences, and these are helping in the understanding of some of the fundamental problems of life. But I think I have derived even more pleasure from the development of the work—seeing the method gradually improving until we were able to read a sequence straight from an autoradiograph. Before this reading became automated, people complained that it was a tedious process, but to me it was always a delight, having in the back of my mind the way we used to do sequences one residue at a time by painstaking partial hydrolysis, fractionation, and analysis. At one stage in the work I would take the autoradiographs home with me and look forward to the pleasure of reading them in the peace of the evening.

At about the same time that the plus and minus method came out, another rapid sequencing technique was developed by Maxam & Gilbert (53, 53a). This was somewhat similar to our method in that it gave an autoradiograph of an ionophoresis from which a sequence could be read off. However, whereas we obtained the mixtures of fragments all terminating at the same nucleotide by a copying procedure using DNA polymerase, they used partial degradation by chemical methods of fragments labeled at their 5' ends. The method worked well and was widely used in preference to the plus and minus method, and in the United States probably in preference to the dideoxy technique until the introduction of the M13 cloning greatly increased the applicability of the dideoxy method. I think it would be fair to say that at present, for a long sequence, the dideoxy-M13 approach is preferable, whereas if one is interested only in a short specific sequence it is probably easier to set up the chemical method.

While the two techniques are to some extent complementary and one must welcome any scientific progress, I cannot pretend that I was altogether overjoyed by the appearance of a competitive method. However, this did not generate any sort of a "rat race," and I do not think it affected our subsequent work at all. I was by no means satisfied with the plus and minus method and certainly the dideoxy method would have been developed in any case. It may be wondered why I did not also try a similar chemical method. I had indeed considered similar ideas but suspect I rejected them because of lack of specificity. Throughout the work on proteins and RNA it had usually been found that enzymic methods were more successful than chemical ones be-

cause of their greater specificity. I was thus rather prejudiced in favor of the use of enzymes and assumed that for a rapid degradative sequencing method to work the reactions would have to be specific. In fact those used by Maxam & Gilbert are not specific, but it turns out that this does not matter and may even result in stronger data than would be obtained by completely specific reactions.

In this account I have dealt almost entirely with our work on methods for sequencing and have said little about the results obtained. This is partly for reasons of space, but also because I consider that my own personal contributions have been to the methods. In the course of the work we have determined many sequences and obtained significant results, which have been reported elsewhere, but in these I have usually been part of a team. I quite enjoyed doing the somewhat routine DNA sequencing work where it was possible to produce results without too much effort, and I probably found this a form of relaxation compared with the more original, exacting, variable, and often frustrating, work on development of methods. With the larger DNA sequences it is necessary to have someone collect and analyze the data, and this was usually done by Bart Barrell, who is something of a wizard with sequences. An exception was when we were working on bacteriophage λ (54), which was sequenced mainly by Alan Coulson and me. At first the routine work of collecting the data was good fun, but I think both of us were happy when the sequence was finally completed. Filling up the last gaps in a sequence can be both slow and frustrating. The main high spots of these sequencing results were two unexpected findings—the overlapping genes in bacteriophage øX174 (55, 56), and the altered genetic code in mitochondrial DNA (57). This work is adequately described in other articles (57–61).

Table 1 summarizes the progress made in sequencing, and lists the various landmarks discussed in this article, starting from Jensen & Evans's positioning of a single amino acid in insulin in 1935 and ending with the recent enormous acceleration in DNA sequencing. At present the number of sequences in the data banks is in the millions, and is still increasing rapidly.

Although I have dealt only with sequencing methods in this article, my interests have not been quite so narrow, and from time to time I have ventured into other fields, although usually unsuccessfully. The most successful diversion was the discovery of the initiator tRNA, formyl-methionyl tRNA (62). This was largely the work of Kjeld Marcker, who came as a visiting worker to the lab at about the time we were first trying to develop methods on RNA. The idea was to get some sequence information near the amino acid binding sites of tRNAs by labeling them with radioactive amino acids. Partial hydrolysis should then yield oligonucleotides of various lengths joined to an amino acid, and methods could be designed for sequencing them. Methionine was chosen to try out the method because it could be labeled at high specific activity with

Table 1 The progress in sequencing

Year	Protein	RNA	DNA	Number of residues	Ref.
1935	Insulin			1	4
1945	Insulin			2	3
1947	Gramicidin S			5	16
1949	Insulin			9	12
1955	Insulin			51	22
1960	Ribonuclease			120	25
1965		$tRNA_{Ala}$		75	32
1967		5S RNA		120	35
1968			Bacteriophage λ	12	45
1978			Bacteriophage φX 174	5,386	61
1981			Mitochondria	16,569	58
1982			Bacteriophage λ	48,502	54
1984			Epstein-Barr virus	172,282	64

^{35}S. To see if the coupling of the methionine to the tRNA had worked, the product was digested with pancreatic ribonuclease and subjected to ionophoresis. The expected adenosyl ester of methionine was identified, but there was another, unexpected, spot present. This could easily have been dismissed as an artifact, but that did not satisfy Marcker. It turned out to be the adenosyl ester of N-formyl methionine, and further experiments enabled him to show that there were two methionyl tRNAs in *Escherichia coli:* one, $tRNA_m^{Met}$, which incorporates methionine within protein chains, and the other, $tRNA_f^{Met}$, which incorporates it only in the N-terminus and is the chain initiator (63).

Unlike many scientists, I decided to retire and give up research when I reached the age of 65. This surprised my colleagues, and to some extent myself also. I had not thought about retirement until I suddenly realized that in a few years I would be 65 and would be entitled to stop work and do some of the things I had always wanted to do and had never had time for. The possibility seemed surprisingly attractive, especially as our work had reached a climax with the DNA sequencing method and I rather felt that to continue would be something of an anticlimax. The decision was I think a wise one—not only because I have greatly enjoyed the new life-style, but also because the aging process was not improving my performance in the laboratory and I think that if I had gone on working I would have found it frustrating and have felt guilty at occupying space that could have been available to a younger person. For more than 40 years I had had wonderful opportunities for research, and had been given the chance to fulfill some of my wildest dreams.

Literature Cited

1. Loewy, A. G., Siekevitz, P. 1969. *Cell Structure and Function,* pp. 193–94. New York: Holt, Rinehart & Winston
1a. Bergmann, M., Niemann, C. 1938. *J. Biol. Chem.* 122:577–96
2. Chibnall, A. C. 1942. *Proc. R. Soc. London Ser. B* 131:136–60
3. Sanger, F. 1945. *Biochem. J.* 39:507–15
4. Jensen, H., Evans, E. A. 1935. *J. Biol. Chem.* 108:1–12
5. Edman, P. 1950. *Acta Chem. Scand.* 4:283–93
6. Gordon, A. H., Martin, A. J. P., Synge, R. L. M. 1943. *Biochem. J.* 37:79–86
7. Gurin, S., Clarke, H. T. 1934. *J. Biol. Chem.* 107:395–99
8. Abderhalden, E., Stix, W. 1923. *Hoppe-Seyler's Z.* 129:143–56
9. Sanger, F. 1947. *Nature* 160:295–96
10. Sanger, F. 1949. *Biochem. J.* 44:126–28
11. Tiselius, A., Sanger, F. 1947. *Nature* 160:433–34
12. Sanger, F. 1949. *Biochem. J.* 45:563–74
13. Sanger, F. 1985. *Curr. Contents* 28(12): 23
14. Gray, W. R., Hartley, B. S. 1963. *Biochem. J.* 89:59
15. Consden, R., Gordon, A. H., Martin, A. J. P. 1944. *Biochem. J.* 38:224–32
16. Consden, R., Gordon, A. H., Martin, A. J. P., Synge, R. L. M. 1947. *Biochem. J.* 41:596–602
17. Sanger, F., Tuppy, H. 1951. *Biochem. J.* 49:463–81
18. Sanger, F., Thompson, E. O. P. 1953. *Biochem. J.* 53:353–66
19. Sanger, F., Tuppy, H. 1951. *Biochem. J.* 49:481–90
20. Sanger, F., Thompson, E. O. P. 1953. *Biochem. J.* 53:366–74
21. Sanger, F., Thompson, E. O. P. 1952. *Biochim. Biophys. Acta* 9:225–26
21a. Ryle, A. P., Sanger, F. 1955. *Biochem. J.* 60:535–40
22. Ryle, A. P., Sanger, F., Smith, L. F., Kitai, R. 1955. *Biochem. J.* 60:541–56
23. Sanger, F. 1959. *Les Prix Nobel en 1958,* pp. 134–46. Reprinted in *Science* 129:1340–45
24. Spackman, D. H., Stein, W. H., Moore, S. 1958. *Anal. Chem.* 30:1190–1206
25. Hirs, C. H. W., Moore, S., Stein, W. H. 1960. *J. Biol. Chem.* 235:633–47
26. Sanger, F., Hocquard, E. 1962. *Biochim. Biophys. Acta* 62:606–7
27. Hartley, B. S., Naughton, M. A., Sanger, F. 1959. *Biochim. Biophys. Acta* 34:243–44
28. Naughton, M. A., Sanger, F., Hartley, B. S., Shaw, D. C. 1960. *Biochem. J.* 77:149–63
29. Sanger, F., Shaw, D. C. 1960. *Nature* 187:872–73
30. Milstein, C., Sanger, F. 1961. *Biochem. J.* 79:456–69
31. Sanger, F. 1963. *Proc. Chem. Soc.* (UK), March 1963:76–83
32. Holley, R. W., Agpar, J., Everett, G. A., Madison, J. T., Marquisee, M., et al. 1965. *Science* 147:1462–65
33. Sanger, F., Brownlee, G. G., Barrell, B. G. 1964. *J. Mol. Biol.* 13:373–98
34. Ling, V. 1972. *J. Mol. Biol.* 64:87–102
35. Brownlee, G. G., Sanger, F., Barrell, B. G. 1968. *J. Mol. Biol.* 34:379–412
36. Adams, J. M., Jeppesen, P. G. N., Sanger, F., Barrell, B. G. 1969. *Nature* 223:1009–14
37. Jeppesen, P. G. N., Barrell, B. G., Sanger, F., Coulson, A. R. 1972. *Biochem. J.* 128:993–1006
38. Murray, K. 1970. *Biochem. J.* 118:831–41
39. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H., Fischer, D. 1973. *Proc. Natl. Acad. Sci. USA* 70:1209–13
40. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H., Fischer, D. 1974. *J. Mol. Biol.* 90:315–33
41. Ziff, E. B., Sedat, J. W., Galibert, F. 1973. *Nature New Biol.* 241:34–37
42. Robertson, H. D., Barrell, B. G., Weith, H. L., Donelson, J. E. 1973. *Nature New Biol.* 241:38–40
43. Billeter, M. A., Dahlberg, J. E., Goodman, H. M., Hindley, J., Weissmann, C. 1969. *Nature* 224:1083–86
44. Wu, R., Kaiser, A. D. 1968. *J. Mol. Biol.* 35:523–27
45. Wu, R., Taylor, E. 1971. *J. Mol. Biol.* 57:491–511
46. Berg, P., Fancher, H., Chamberlin, M. 1963. In *Symposium on Informational Macromolecules,* ed. H. Vogel, V. Bryson, J. O. Lampen, pp. 467–83. New York/London: Academic
47. Sanger, F., Coulson, A. R. 1975. *J. Mol. Biol.* 94:441–48
48. Atkinson, M. R., Deutscher, M. P., Kornberg, A., Russell, A. F., Moffatt, J. G. 1969. *Biochemistry* 8:4897–904
49. Sanger, F., Nicklen, S., Coulson, A. R. 1977. *Proc. Natl. Acad. Sci. USA* 74:5463–67
50. Gronenborn, B., Messing, J. 1978. *Nature* 272:375–77
51. Messing, J., Vierira, J. 1982. *Gene* 19: 269–76
52. Deininger, P. 1983. *Anal. Biochem.* 129:216–23
53. Maxam, A. M., Gilbert, W. 1977. *Proc. Natl. Acad. Sci. USA* 74:560–64

53a. Maxam, A. M., Gilbert, W. 1980. *Methods Enzymol.* 65(P. 1):499–560
54. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., Petersen, G. B. 1982. *J. Mol. Biol.* 162:729–73
55. Barrell, B. G., Air, G. M., Hutchison, C. A. III. 1976. *Nature* 264:34-41
56. Smith, M., Brown, N. L., Air, G. M., Barrell, B. G., Coulson, A. R., et al. 1977. *Nature* 265:702–5
57. Barrell, B. G., Anderson, S., Bankier, A. T., de Bruijn, M., Chen, E., et al. 1980. *Proc. Natl. Acad. Sci. USA* 77:3164–66
58. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., et al. 1981. *Nature* 290:457–65
59. Anderson, S., de Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F., et al. 1982. *J. Mol. Biol.* 156:683–717
60. Sanger, F. 1981. *Les Prix Nobel en 1980*, pp. 143–59. Reprinted in *Biosci. Rep.* 1:3–18; *Science* 214:1205–10
61. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., et al. 1978. *J. Mol. Biol.* 125:225–46
62. Marcker, K., Sanger, F. 1964. *J. Mol. Biol.* 8:835–40
63. Clark, B. F. C., Marcker, K. A. 1966. *J. Mol. Biol.* 17:394–406
64. Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., et al. 1984. *Nature* 310:207–11