

# Perspectives

## Anecdotal, Historical and Critical Commentaries on Genetics

*Edited by James F. Crow and William F. Dove*

### The First Sequence: Fred Sanger and Insulin

**Antony O. W. Stretton**

*Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706*

**F**RED Sanger is an amazingly modest man, and his own retrospective, written after he retired, a delightful prefatory chapter for the Annual Reviews of Biochemistry, is called "Sequences, sequences, and sequences" (SANGER 1988). In it he describes the paths that led to the successful methods he developed for the sequencing of proteins, then RNA, and then DNA. What a career!

Especially now, with the human genome largely finished, it is almost impossible to imagine a world without sequences of proteins and of nucleic acids. The fact that it has been only 50 years since Sanger showed that there was such a thing as the unique amino acid sequence of a protein seems amazing from where we now stand—sequences are such a dominant part of the world we work in today. Amazing, but perhaps not surprising, when we think of, say, the change in the power of computers that has occurred over just about the same time period. (I would love to be around, marbles intact, 50 years from now, to find out how the brain really works.)

Before Fred's work, it was already known that different proteins had different amino acid compositions, different biological activities, and different physical properties and that genes had an important role in controlling them. But in a world of biochemistry dominated by the role of enzymes in intermediary metabolism, it was not at all clear how molecules as large as proteins could be synthesized; the idea that proteins were stochastic molecules, with a sort of "center of gravity" of structure but with appreciable microheterogeneity, was taken seriously. This is the paradigm that Fred's results shifted.

This essay is a celebration of his first triumph: the first complete amino acid sequence determination of a protein, the B chain of insulin, published just over 50 years ago. Although I was not involved in the work in any way, I was scientifically aware enough to feel its contemporary impact. In 1957, Vernon Ingram, who at the time was working in the MRC Unit for the Study of the Molecular Structure of Biological Systems in the Cavendish Laboratory at Cambridge (later this morphed

into the Laboratory of Molecular Biology), took me on as a research student. Vernon had just shown that sickle-cell hemoglobin differed from normal hemoglobin by a single amino acid substitution, the first characterization of the molecular consequences of mutation on proteins (INGRAM 1956, 1957); earlier NEEL (1949) had shown that sickle-cell anemia is inherited as a Mendelian character, and PAULING *et al.* (1949) showed that sickle-cell hemoglobin differed electrophoretically from normal hemoglobin and coined the term "molecular disease." I had been trained as a chemist and knew nothing about proteins: I had heard Alex Todd's exciting lectures to chemistry students on the organic chemistry of natural products that included vitamins, steroids, and nucleic acids, but not a word about proteins! To work on hemoglobin in Vernon's laboratory, I was going to have to learn to become a protein chemist, and Francis Crick urged me to go and see Fred Sanger, who was at that time in the Biochemistry Department. In preparation for this, I read Fred's 1952 review in *Advances in Protein Chemistry* entitled "The arrangement of amino acids in proteins" (SANGER 1952). I had also read a preprint of Francis's own review, "On protein synthesis" (CRICK 1958). Each review was completely stunning, but in very different ways. Francis's article was full of elegant theory, with great leaps of speculation (at the time the existence of messenger RNA and transfer RNA was not known), but conveying a sense of joy that was as thrilling as that emerging from the *Origin of Species*. Fred's was full of technique, and his review was scholarly, with major emphasis on experiments. Both of them made a huge impression on me, but I was particularly struck by Fred's emphasis on the importance of developing new techniques. Today, when "hypothesis-driven research" has somehow become a gold standard, in many areas one must be very brave to admit to working on techniques, at least in the biological sciences. I think it may be different in chemistry—at least that is what my recent contacts with mass spectrometry suggest. Bill Dove has noted that Al Hershey (Nobel Prize winner in 1969), another temperamentally modest scientist who made

seminal contributions, was also passionate about methods: "There's nothing like technical progress! Ideas come and go, but technical progress cannot be taken away" (Hershey, quoted in DOVE 1987).

Another profound lesson I learned was that completely different scientific styles can be equally successful and valid and that personality has little to do with success. Again, the contrast between Fred and Francis is vast. After as little as half a minute with Francis, you know he has an exceptional mind, but with Fred it is much more subtle. He was brought up as a Quaker, which probably has a lot to do with his low-key and quiet demeanor, and during World War II he was a conscientious objector. In his 1988 article he writes, "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" and "Unlike most of my scientific colleagues, I was not academically brilliant," and this after he had won two Nobel Prizes! So much for academic brilliance.

Another unusual aspect of Fred's character is his ability to pick and nurture people. I first got to know Fred indirectly through his graduate student Mike Naughton. Vernon moved to MIT in 1958 and took me with him, and Mike joined Vernon's lab as a post-doc during my second year there. Mike and I became good friends, and I learned a lot of Fred's techniques from Mike, so in some way I feel like one of Fred's scientific grandchildren. Mike is a big, gentle Irishman from the West, who loved to sing and tell awful jokes. He had been a schoolteacher, and after doing his National Service in the Royal Air Force, he joined Fred as a technician. Soon he was transformed into Fred's Ph.D. student, working also with Brian Hartley on a beautiful piece of work showing that the sequences around the active sites of the pancreatic serine proteases are identical (HARTLEY *et al.* 1959). Fred has done the same thing for at least two other people that I know, each taken on as a technician and turned into a Ph.D.—Bart Barrell, who is one of the world's best nucleic acids sequencers, and Alan Coulson, who later became John Sulston's right-hand man on the genome projects of both Worm and Human.

I will quote extensively from the introduction to Fred's 1952 review, because it sets the stage beautifully for what was happening, a reevaluation of the nature of proteins. Fred wrote:

It has frequently been suggested that proteins may not be pure chemical entities but may consist of mixtures of closely related substances with no absolute unique structure. The chemical results obtained so far suggest that this is not the case, and that a protein is really a single chemical substance, each molecule of one protein being identical to every other molecule of the same pure protein.

Another earlier model (BERGMANN and NIEMANN

1938) had suggested that proteins had periodic arrangements of amino acids, but the sequence of insulin ruled that out too. He added,

These results [*the insulin sequence*] would imply an absolute specificity for the mechanisms responsible for protein synthesis and this should be taken into account when considering such mechanisms.

He concludes,

It is certain that proteins are extremely complex molecules but they are no longer completely beyond the reach of the chemist, so that we may expect to see in the near future considerable advances in our knowledge of the chemistry of these substances which are the essence of living matter.

Remember that this was written only one year after Fred and Hans Tuppy had solved the structure of the B chain of insulin. The accuracy of protein synthesis remained an issue for many years after that.

The previous review of the covalent structure of proteins had been written by Syngé in 1943 (SYNGE 1943). In 1952, Fred writes:

... up to that time [1943] only a few simple peptides had been clearly identified from proteins by the classical and rather laborious methods of organic chemistry and Syngé concluded that "the main obstacle to progress in the study of protein structure by the methods of organic chemistry is inadequacy of technique!" Probably the greatest advance that has been made recently in this field was the development by MARTIN and SYNGE (1941) of the entirely new technique of partition chromatography. The great problem in peptide chemistry has always been to find methods of fractionating the extremely complex mixtures produced by the partial degradation of a protein. Older methods of fractional crystallization and precipitation with various reagents were as a rule inadequate to deal with these mixtures, and countercurrent methods of high resolving power, which could fractionate non-volatile, water-soluble substances, were needed. Partition chromatography, especially in the form of paper chromatography (CONSDEN *et al.* 1944), is such a method, so that it has already been possible to identify as breakdown products of proteins more peptides using this technique than had previously been identified by the classical methods of organic chemistry. During the last few years, work in this field has centered largely on the development of methods, so that this review will be more a consideration of techniques and their uses than a discussion of results, which are still rather few.

**N-terminal sequences of insulin:** One of the main reasons Sanger chose insulin for this work is that it was one of the few proteins available in pure form, and it was available in gram quantities because of its medical importance. At the time, the physical chemical evidence suggested a molecular weight of about 12,000. Fred invented the N-terminal labeling method using 1:2:4 fluorodinitrobenzene (FDNB), which reacts with amino groups under mild conditions that avoid degradation of the polypeptide chain. After complete acid hydrolysis of the dinitrophenyl (DNP)-protein, the DNP groups remain attached to the N-terminal amino acid and can

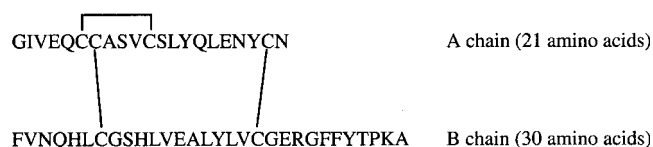


FIGURE 1.—The structure of bovine insulin.

be isolated and identified. Fred showed that there were four N-terminal residues per 12K insulin molecule, two of which were glycine and two phenylalanine (SANGER 1945), suggesting that there were four polypeptide chains in the 12K molecule. Cysteine was present, so it was thought that the chains were held together by -S-S- bridges, and indeed after performic acid oxidation, which splits the -S-S- bridges, insulin could be fractionated by precipitation into an A fraction and a B fraction; the A fraction had N-terminal glycine, and the B fraction had phenylalanine (SANGER 1949). The two fractions had different amino acid compositions, and neither contained tryptophan. Later it became clear that the 12K molecule is a noncovalent dimer of the fundamental molecular unit (HARFENIST and CRAIG 1952), comprising one A chain and one B chain (see Figure 1).

The lack of tryptophan was particularly fortunate, because it degrades upon acid hydrolysis, and one of the most important methods Fred used to get at the structure of insulin, with great success, was partial acid hydrolysis, which splits the peptide bonds almost randomly (more about that later). In fact, the first amino acid sequences of insulin came from partial acid hydrolysis of DNP-labeled A and B fractions. Some DNP peptides can be extracted into ethyl acetate from acid solution and then separated by silica gel columns; since DNP compounds are usually yellow, this was real "chromatography." For the B fraction, these peptides turned out to be the DNP-labeled N-terminal Phe followed by one or more other amino acids. Fred identified the DNP-amino acid and the other amino acids in the peptide after complete acid hydrolysis and then assembled the N-terminal sequence Phe.Val.Asp.Glu. Among the peptides that contained Asp or Glu, several different peptides had the same amino acid composition, so he concluded that both Asp and Glu were amidated in the original sequence.

Other DNP peptides were not extracted into ethyl acetate. They were peptides derived from internal sequences surrounding Lys, to which DNP was linked by the amino group on the side chain. Since they all had free N-terminal amino groups (liberated from internal peptide bonds by partial acid hydrolysis), they were positively charged in acid, which explains why they did not extract into organic solvents. Amino acid analysis and relabeling with FDNB to determine the end groups gave the internal sequence Thr.Pro.Lys.Ala. The A fraction yielded the N-terminal sequence Gly.Ileu.Val.Glu.Glu.

This article (SANGER 1949) was pivotal—it showed for the first time that at least some of the amino acids were in a unique sequence in insulin. Furthermore, the A and B fractions each yielded a unique sequence, suggesting that there were only two, not four, species of peptide chain in insulin—an A chain that contained about 20 amino acids and a B chain with about 30 amino acids—and he already had the sequence of over a quarter of the B chain! Even more important, this article showed that it should be possible in principle to determine the whole structure of each chain simply by extending the methods developed in this article—partial hydrolysis, fractionation of the products, end group analysis, and further partial hydrolysis of the longer products. In practice, it is the fractionation methods that are limiting, and the complexity of the mixture has to be controlled to match them. SANGER and TUPPY (1951a) did many experiments to approach a compromise between the ideal and the feasible, which meant concentrating on the later stages of the hydrolysis, where the average size of the peptides, and therefore also their number in the mixture, were relatively low.

**The complete sequence of the B chain:** The B chain was tackled first (SANGER and TUPPY 1951a). Partial acid hydrolysis of the whole untagged chain yielded many more products to be separated, and the problem of separation of pure peptides from such a complex mixture was severe. They used several methods. To fractionate acidic peptides, they used batch absorption on ion exchange resins; they adjusted the pH so that the cysteic acid peptides (from the performic oxidation of cystine) could be separated from the aspartate and glutamate peptides, taking advantage of the very low pK of the resulting sulfonic acid. After charcoal fractionation, aimed at isolating peptides containing aromatic amino acids, they used batch ionophoresis in solution with several in-series compartments containing acids or bases to modify the charge on peptides (the compartments were separated by membranes that sound pretty exotic—formolized gelatin, cellophan; Syngé had used formolized sheepskin parchment). These relatively crude fractions, which contained between 8 and 25 peptides, were then subjected to two-dimensional paper chromatography, and the peptides were detected (usually) by lightly staining with ninhydrin. Finally, they were able to characterize 23 dipeptides, 15 tripeptides, 9 tetrapeptides, 2 pentapeptides, and 1 hexapeptide by amino acid composition and end-group analysis. End-group analysis was subtractive, since at this stage no sensitive and reliable method for identifying all the individual DNP-amino acids had been developed. Then the exciting bit—and you can tell from the article that they loved this—was the assembly into a longer sequence. One early discovery was that the peptide bonds N terminal to Ser or Thr residues are particularly labile to acid hydrolysis and always cleave first, so they found no dipeptides with C-terminal Ser or Thr. This prevented the



complete assembly, but they deduced the sequence of five fragments of the B chain: two tetrapeptides (Thr. Pro.Lys.Ala and Gly.Glu.Arg.Gly), a pentapeptide (Tyr. Leu.Val.Cys.Gly), a hexapeptide (Ser.His.Leu.Val.Glu. Ala), and an octapeptide (Phe.Val.Asp.Glu.His.Leu.Cys. Gly, which included the N terminus).

The completion of the sequence depended on the use of specific proteases—trypsin, chymotrypsin, and pepsin—to produce larger fragments (SANGER and TUPPY 1951b). At the time this was considered very risky, because it was believed that proteases could catalyze the synthesis as well as the cleavage of peptide bonds. Happily, the experiments on insulin proved that this was not true. Previous work by BERGMANN and his colleagues (reviewed in BERGMANN and FRUTON 1941), using simple synthetic substrates, had shown that these enzymes cleave at different residues, trypsin at lysine or arginine and chymotrypsin preferring aromatic amino acids, and the peptides obtained from proteolysis of insulin confirmed this. Pepsin was less predictable, but tended to cleave around aromatic amino acids and leucine, but also at several other sites. With protease specificity, the proof of the pudding is in the protein substrate, and the experiments on the cleavage of insulin were the first real test.

The main advantage of protease digestion is that trypsin and chymotrypsin cleave at few sites, but do so relatively completely, so the complexity of the mixture to be separated is low. The protein is chopped into neat chunks that can be isolated and characterized by the same methods as in SANGER and TUPPY (1951a); in addition, some of the peptides from one digest were further digested with a second enzyme, to identify overlapping regions in the different series of peptides. This time they obtained enough overlapping sequences (in particular, sequences spanning Ser and Thr residues) to finish the assembly, and the sequence of the 30 amino acids in the B chain was completed. These techniques, especially the generation of overlapping sets of peptides derived from different protease digests, became the standard method for protein sequence determination for many years.

Fred's initial approach to sequencing by random cleavage was too difficult for larger proteins, mainly because of the problems of fractionation of complex mixtures. At the time this work was done, the Edman sequential degradation technique had already been described (EDMAN 1950). This method removes amino acids sequentially from the N terminus; it later was the method of choice for the direct sequencing of peptides and proteins and became highly automated and sensitive at the sub-picomole level. In his memoir, Fred says that he did not use it because the products were not colored like the yellow DNP-derivatives, and so fractionations were difficult to follow in the absence of flow spectrophotometers and reliable fraction collectors.

The DNP compounds could be seen as yellow bands moving down the column.

**The A chain:** Two years later, SANGER and THOMPSON (1953a,b) completed the sequence of the A chain, which, although shorter (21 amino acids compared with 30 in the B chain), was technically more difficult. Again, they used partial acid hydrolysis of the whole chain and then protease digestion to give larger blocks. In these articles a new fractionation method was used, ionophoresis in silica gels, a tricky method that was very useful for these experiments but was soon outdated and replaced by the powerful paper ionophoresis. Chromatographic methods for identifying the individual DNP amino acids were developed, so now they could positively identify the N termini of peptides. These partial acid fragments could be assembled into longer sequences: an octapeptide that contained the known N-terminal sequence, a nonapeptide, a tripeptide, and a dipeptide, which together included all the amino acids in the A chain, but again the lability of the peptide bond N terminal to serine and threonine prevented the final assembly. Peptic and chymotryptic proteolysis produced the appropriate fragments, and the sequence was done, with C-terminal asparagine confirmed by carboxypeptidase A digestion. They conclude this account:

It would thus seem that no general conclusions can be drawn from these results concerning the general principles which govern the arrangement of the amino-acid residues in protein chains. In fact, it would seem more probable that there are no such principles, but that each protein has its own unique arrangement; an arrangement which endows it with its particular properties and specificities and fits it for the function that it performs in nature.

Yes! This is the conclusion that is so monumental, and had so much influence on the rise of molecular biology (*sine qua non*).

**The -S-S- bonds:** The A chain and B chain by themselves are inactive physiologically. Sanger and his colleagues Ryle, Smith, and Kitai went on to finish another functionally very important piece of covalent chemistry: the three disulfide bonds in the unoxidized molecule. They had to develop new methods, because the disulfide bonds rearrange under some of the conditions they used to break peptide bonds. They discovered that adding thiol blocking reagents like N-ethylmaleimide prevented the exchange under the slightly alkaline conditions used for digestion with pancreatic proteases. One of the interchain disulfide bonds was easily identified, but the A chain includes two adjacent cysteines, and they found no protease that could cleave the peptide bond between them. Further experiments showed that rearrangement under acid conditions was inhibited by added thiols, so they were able to proceed with their original sequencing techniques by partial acid hydrolysis. However, the separation of partial acid hydrolysis products was now even more challenging because the

mixtures were inevitably more complex (two different random cleavage products joined together by the disulfide bridges). By now, paper ionophoresis had been added to the fractionation methods, and, by using different pH conditions and combining the electrophoretic separations with paper chromatography in various solvent systems, they found the necessary fragments that gave the unambiguous assignments of the disulfide bonds (RYLE *et al.* 1955)—there were two interchain bonds, and one intrachain bond in the A chain (see Figure 1). The Nobel committee was swift to recognize the importance of this work, and Fred's first Prize came in 1958 (his second Prize, awarded in 1980, was for sequencing nucleic acids).

Two overall comments: first, the methods used in this sequence determination were largely nonquantitative. Amino acid analysis was done by comparing the ninhydrin staining intensities by eye with standards, and with short peptides that is good enough. A little later, Moore and Stein and their colleagues developed methods for separating peptides on ion exchange columns (HIRS *et al.* 1960) and for quantitative amino acid analysis, a procedure that was soon semi-automated (SPACKMAN *et al.* 1958). My own sequencing, on human hemoglobin A<sub>2</sub> and bacteriophage T4 head protein, was hybrid; I isolated peptides mostly by paper ionophoresis and chromatography, but used quantitative amino acid analysis to determine the compositions. After a long hiatus, I have recently returned to peptide chemistry, and it is a different world. Peptide isolation is now done by high performance liquid chromatography (HPLC), which has excellent resolution and recovery; sequencing is totally automated and at least three orders of magnitude more sensitive, and in most cases it is done by someone else, using a big, expensive machine in one or another "Biotech Center"—a lot of the fun is lost. Nowadays, mass spectrometry is of growing importance in sequencing. The peptides can be randomly cleaved *in vacuo* and the fragments characterized by molecular mass to assemble the sequence. Same idea as partial acid hydrolysis! One serious problem is that leucine and isoleucine have the same mass. So do lysine and glutamine, but they are chemically different, and it is easy to identify lysine by the mass shift after acetylation. The biggest change is that most protein sequencing is done indirectly from DNA sequences, and only those of us who work on post-translationally processed proteins and peptides (neuropeptides in my case) are still forced to work with the amino acids.

Second, on a very different level, those of us who did sequencing using the original methods would probably make a really interesting epidemiological group to study, because we were heavily exposed to a variety of organic solvents in the separations on paper, both by ionophoresis and chromatography. For ionophoresis, the paper was immersed in toluene, which acted as a coolant (it was cooled with circulating water in coils

immersed in the toluene), and it was hard to avoid getting it on your hands as you manipulated the paper. Later, toluene was replaced by Varsol, a refined, high-boiling-point petroleum fraction. The most commonly used buffers contained pyridine and were really stinky. The mixture to be separated was loaded onto the paper in a small volume and dried with a hair dryer; then the buffer was applied to the dry paper, which was spread out on a glass plate. The art was in wetting the paper on each side so that the solvent fronts met at the origin at exactly the same time, and this needed a steady hand and was best not done right after coffee or tea time. This concentrated the sample into a very narrow line, if you did it right, but also left you breathing pyridine for quite a while. The solvents for paper chromatography were also pretty pungent, and on the occasions (thankfully rare) when I used hexane/formic acid, my nose would bleed, and I would have a headache until the next day. In those days we were quite ignorant of the dangers of this sort of exposure.

It is also fascinating to see the evolution of the methodology in the series of articles that describe the covalent structure of insulin. We tend to recognize Fred's development of the FDNB method of N-terminal labeling as crucial, and indeed it was, but his exploration of chemical and enzymatic cleavage methods and his use of new fractionation methods for mixtures of peptides, were also critical for his success. Paper chromatography was a relatively new technique that he and his colleagues used extensively, but even two-dimensional chromatography was not adequate for the complete separation of the complex mixtures generated by partial acid hydrolysis, so they used various prefractionation methods, including absorption onto charcoal to selectively remove peptides containing aromatic amino acids, batch ionophoresis in solution or in silica gels, and ion exchange. Relatively late in the game (RYLE *et al.* 1955) they started to use paper ionophoresis, a technique that separated peptides almost exclusively on the basis of charge and molecular mass. It was even more effective when used in tandem with chromatography, which was also sensitive to hydrophobicity [this technique was the basis of "fingerprinting," Vernon INGRAM's (1956) method for comparing normal and sickle-cell hemoglobin].

Fred Sanger's stunning, startling, mind-expanding 1951 articles (with Hans Tuppy) on the sequence of the B chain of insulin deserve a huge worldwide Jubilee celebration, particularly among geneticists! The linearity of genetic maps was already well known, and a few years later Seymour BENZER (1955) showed that multiple sites of mutation within a single gene also mapped onto a strictly linear construct. Fred had revealed another linear world, that of polypeptides. The molecular biologists put the two together, and the fact that there was a well-defined sequence in proteins led to the thought that there had to be a genetic code. Later, Fred came back and sequenced the other two players (RNA

and DNA) in the genetic control of protein structure, but that is another story.

I am very grateful to Philippa Claude for her insightful criticisms and suggestions; as always, her judgment is penetrating and accurate.

#### LITERATURE CITED

- BENZER, S., 1955 Fine structure of a genetic region in bacteriophage. *Proc. Natl. Acad. Sci. USA* **41**: 344-354.
- BERGMANN, M., and C. NIEMANN, 1938 On the structure of silk fibroin. *J. Biol. Chem.* **122**: 577-596.
- BERGMANN, M., and J. S. FRUTON, 1941 The specificity of proteinases. *Adv. Enzymol.* **1**: 63-98.
- CONSDEN, R., A. H. GORDON and R. L. M. SYNGE, 1944 Qualitative analysis of proteins: a partition chromatographic method using paper. *Biochem. J.* **38**: 224-232.
- CRICK, F. H. C., 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12**: 138-163.
- DOVE, W. F., 1987 Paradox found. *Genetics* **115**: 217-218.
- EDMAN, P., 1950 Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **4**: 283-293.
- HARFENIST, E. J., and L. C. CRAIG, 1952 The molecular weight of insulin. *J. Am. Chem. Soc.* **74**: 3087-3089.
- HARTLEY, B. S., M. A. NAUGHTON and F. SANGER, 1959 The amino acid sequence around the reactive serine of elastase. *Biochim. Biophys. Acta* **34**: 243-244.
- HIRS, C. H. W., S. MOORE and W. H. STEIN, 1960 The sequence of the amino acid residues in performic acid-oxidized ribonuclease. *J. Biol. Chem.* **235**: 633-647.
- INGRAM, V. M., 1956 A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* **178**: 792-794.
- INGRAM, V. M., 1957 Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* **180**: 326-328.
- MARTIN, A. J. P., and R. L. M. SYNGE, 1941 A new form of chromatography employing two liquid phases. *Biochem. J.* **35**: 1358-1368.
- NEEL, J. V., 1949 The inheritance of sickle cell anemia. *Science* **110**: 64-66.
- PAULING, L., H. A. ITANO, S. J. SINGER and I. C. WELLS, 1949 Sickle cell anemia, a molecular disease. *Science* **110**: 543-548.
- RYLE, A. P., F. SANGER, L. F. SMITH and R. KITAI, 1955 The disulphide bonds of insulin. *Biochem. J.* **60**: 542-556.
- SANGER, F., 1945 The free amino groups of insulin. *Biochem. J.* **39**: 507-515.
- SANGER, F., 1949 The terminal peptides of insulin. *Biochem. J.* **45**: 563-574.
- SANGER, F., 1952 The arrangement of amino acids in proteins. *Adv. Protein Chem.* **7**: 1-66.
- SANGER, F., 1988 Sequences, sequences, and sequences. *Annu. Rev. Biochem.* **57**: 1-28.
- SANGER, F., and E. O. P. THOMPSON, 1953a The amino-acid sequence in the glycyl chain of insulin. 1. The investigation of lower peptides from partial hydrolysates. *Biochem. J.* **53**: 353-366.
- SANGER, F., and E. O. P. THOMPSON, 1953b The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* **53**: 366-374.
- SANGER, F., and H. TUPPY, 1951a The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochem. J.* **49**: 463-481.
- SANGER, F., and H. TUPPY, 1951b The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* **49**: 481-490.
- SPACKMAN, D. H., W. H. STEIN and S. MOORE, 1958 Automatic recording apparatus for use in the chromatography of amino acids. *Anal. Chem.* **30**: 1190-1209.
- SYNGE, R. L. M., 1943 Partial hydrolysis products derived from proteins and their significance for protein structure. *Chem. Rev.* **32**: 135-172.