CLASSIC PAPER

FREDERICK SANGER

The chemistry of insulin

Nobel Lecture, December 11, 1958

BY:
ANUBHAV MAHAPATRA
13-01-2024

NOBEL LECTURE

It is great pleasure and privilege for me to give an account of my work on protein structure and I am deeply sensitive of the great honour that has been done to me in recognizing my work in this way. Since the work on insulin has extended over about 12 years it will be necessary to give a somewhat simplified account and to omit most of the work that did not contribute directly to the main problem, the determination of the structure of a protein.

In 1943 the basic principles of protein chemistry were firmly established. It was known that all proteins were built up from amino acid residues bound together by peptide bonds to form long polypeptide chains. Twenty different amino acids are found in most mammalian proteins and by analytical procedures it was possible to say with reasonable accuracy how many residues of each one was present in a given protein. Practically nothing, however, was known about the relative order in which these residues were arranged in the molecules. This order seemed to be of particular importance, since although all proteins contained approximately the same amino acids they differed markedly in both physical and biological properties. It was thus concluded that these differences were dependent on the different arrangement of the amino-acid residues in the molecules. Although very little was known about amino-acid sequence, there was much speculation in this field.

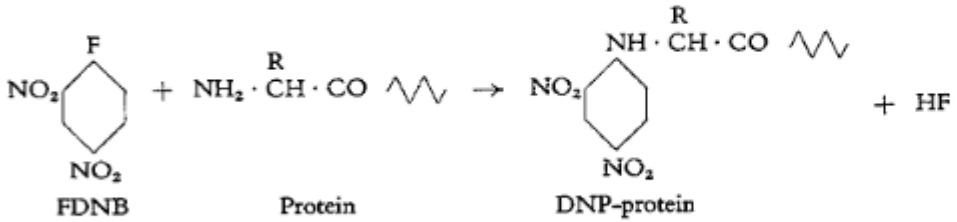
Due largely to the work of Chibnall and his colleagues* insulin had been studied in considerable detail. It had a somewhat simpler composition than most proteins, in that two of the commonly occurring amino acids, tryptophan and methionine were absent and an accurate analysis was available.

Moreover, using the Van Slyke procedure, Chibnall had shown that insulin was peculiar in having a high content of free α-amino groups. This indicated that it was composed of relatively short polypeptide chains since free α-amino groups would be found only on those residues (the N-terminal residues) which were present at one end of a chain. Thus the number of chains could be determined from the number of these N-terminal residues. The nature of one of these N-terminal residues was in fact known. Jensen & Evans[2] had shown that the phenylhydantoin of phenylalanine could be isolated from an acid hydrolysate of insulin that had been treated with phenylisocyanate, thus indicating that phenylalanine was at the end of one of the chains. At that time this was the only case where the position of an amino acid in a protein was known.

CONTD.

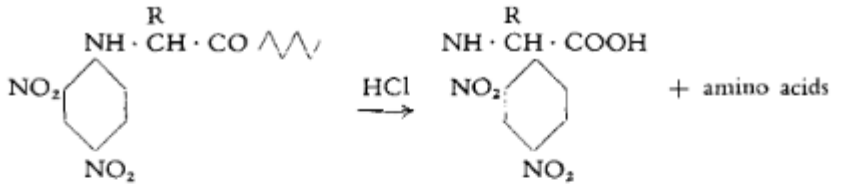
There was considerable doubt about the actual molecular weight of insulin and hence the number of amino acid residues present. Values varying from 36,000 to 48,000 were reported by physical methods but it was shown by Gutfreund[3] that these high values were due to aggregation and it was suggested that the real molecular weight or subunit was 12,000. This indicated that there were about 100 residues in the molecule. More recently Harfenist & Craig[4] have shown that the actual value is about 6,000; however during most of our work it was believed to be 12,000.

In order to study in more detail the free amino groups of insulin and other proteins, a general method for labelling them was worked out[5]. This was the dinitrophenyl (or DNP) method. The reagent used was 1:2:4 fluoro-dinitrobenzene (FDNB) which reacts with the free amino groups of a protein or peptide to form a DNP derivative:



FDNB Protein DNP-protein

The reaction takes place under mildly alkaline conditions which normally do not cause any breakage of the peptide bonds.



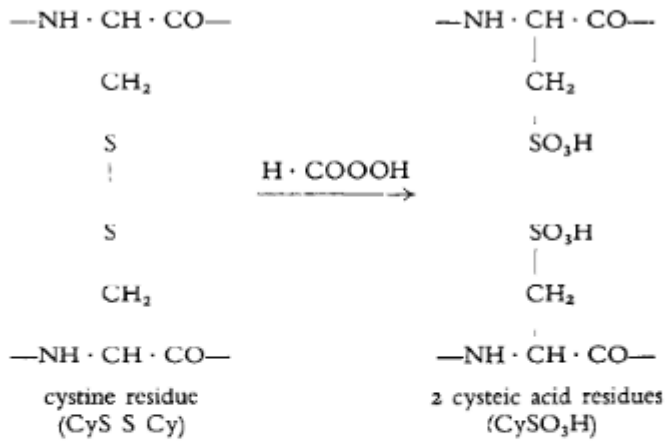
DNP-protein DNP-amino-acid

The DNP-amino acids are bright yellow substances and can be separated from the unsubstituted amino acids by extraction with ether. They could be fractionated by partition chromatography, a method which had just been introduced by Gordon, Martin & Synge[6] at that time. The DNP-amino acids could then be identified by comparison of their chromatographic rates with those of synthetic DNP-derivatives. In the original work on insulin, silica-gel chromatography was used, though more recently other systems, particularly paper chromatography, have been found more satisfactory. Having separated and identified the DNP-derivatives they could be estimated calorimetrically.

It seemed probable that the chains of insulin were joined together by the disulphide bridges of cystine residues. Insulin is relatively rich in cystine and this was the only type of cross-linkage that was definitely known to occur in proteins. It was thus next attempted to separate the peptide chains by splitting the disulphide bridges. Earlier attempts to do this by reduction to -SH derivatives had not proved successful and had given rise to insoluble products 'which were probably due to some type of polymerization. More satisfactory results were obtained by oxidation with performic acid[7]. The

CONTD.

cvstine residues were converted to cysteic acid residues thus breaking the cross-links:



Performic acid also reacts with methionine and tryptophan residues, the two amino acids which fortunately were absent from insulin.

From the oxidized insulin two fractions could be separated by precipitation methods. One (fraction A) contained glycine and the other (fraction B) phenylalanine N-terminal residues. Fraction A was acidic and had a simpler composition than insulin, in that the six amino acids: lysine, arginine, histidine, phenylaline, threonine, and proline, were absent from it. It thus had no basic amino acids, which were found only in fraction B.

Table 1.

Peptide	*Products of complete hydrolysis of peptide*	*Products of partial hydrolysis*	*Structure*	*Yield from DNP* insulin*
B1	DNP-phenylalanine	—	DNP-Phe	13
B2	DNP-phenylalanine Valine	B1	DNP-Phe · Val	16
B3	DNP-phenylalanine Valine Aspartic acid	B1, B2	DNP-Phe · Val · Asp	13
B4	DNP-phenylalanine Valine Aspartic acid Glutamic acid	B1, B2, B3	DNP-Phe · Val · Asp · Glu	30
Other bands giving B4 on partial hydrolysis				20
			Total	92

* Moles peptide as per cent of total N-terminal phenylalanine residues of insulin.

These results, besides giving information about the position of certain residues in the polypeptide chains, showed for the first time that the molecule was composed of only two types of chains and that if the molecular weight was 12,000 as was then believed, then the molecule was built up of two identical halves.

Proteolytic enzymes are much more specific than is acid since only a few of the peptide bonds are susceptible. They give rise to larger peptides which in general are more difficult to fractionate by paper chromatography. However there are relatively few of them so that the mixtures are less complex. In this initial work we used essentially the same methods for studying the

CONTD.

enzymic peptides as we had used for the acid ones, depending largely on paper chromatography for fractionation, although more recently it has been shown that better separations can be obtained by ion-exchange chromatography and by ionophoresis.

As an example we may consider a peptide Bp3 obtained by the action of pepsin. It had the following composition Phe (CySO,H, Asp, Glu, Ser, Gly, Val, Leu, His) of which the most important components are aspartic acid and serine since they occur only once in the chain. Aspartic acid is present only in the N-terminal sequence 1 and serine is in sequence 5. This shows that all of sequence I and at least the N-terminal part of sequence 5 is present in peptide Bp3. That none of the other sequences are present follows from the fact that Bp3 contains no arginine (sequence 2), threonine, proline, or lysine (sequence 3) or tyrosine (sequence 4). One may thus conclude that the two sequences are joined together. By studying other peptides obtained by the action of pepsin, trypsin and chymotrypsin it was possible to find out how the various sequences were arranged and to deduce the complete sequence of the phenylalanyl chain which is shown below:

Phe ·Val ·Asp ·Glu ·His ·Leu ·$CySO_3H$ ·Gly ·Ser ·His ·Leu ·Val ·Glu ·Ala ·Leu ·Tyr ·Leu ·Val ·$CySO_3H$ ·Gly ·Glu ·Arg ·Gly ·Phe ·Phe ·Tyr ·Thr ·Pro ·Lys ·Ala.

Essentially similar methods were used to determine the sequence of fraction A[12]. Although the shorter of the two chains, the determination of its structure was more difficult. Fraction B contains several residues that occur only once in the molecule and this helps considerably in interpreting the results, whereas fraction A has only a few such residues and these are all near one end. Also fraction A is much less susceptible to enzymic hydrolysis. It is not attacked by trypsin and there is a sequence of thirteen residues which is not split by chymotrypsin or pepsin either.

The sequence of fraction A was found to be:

Gly ·Ileu ·Val ·Glu ·Glu ·$CySO_3H$ ·$CySO_3H$ ·Ala ·Ser ·Val ·$CySO_3H$ ·Ser ·Leu ·Tyr ·Glu ·Leu ·Glu ·Asp ·Tyr ·$CySO_3H$ ·Asp.

Having determined the structure of the two chains of insulin the only remaining problem was to find how the disulphide bridges were arranged. About this time it was shown by Harfenist & Craig that the molecular weight of insulin was of the order of 6,000, so that it consisted of two chains containing three disulphide bridges, and not of four chains as we had originally thought. The fact that fraction A contained four cysteic acid residues whereas fraction B had only two indicated that two bridges must connect the two chains together and one must form an intrachain bridge connecting one part of the A chain with another part of the same chain.

CONTD.

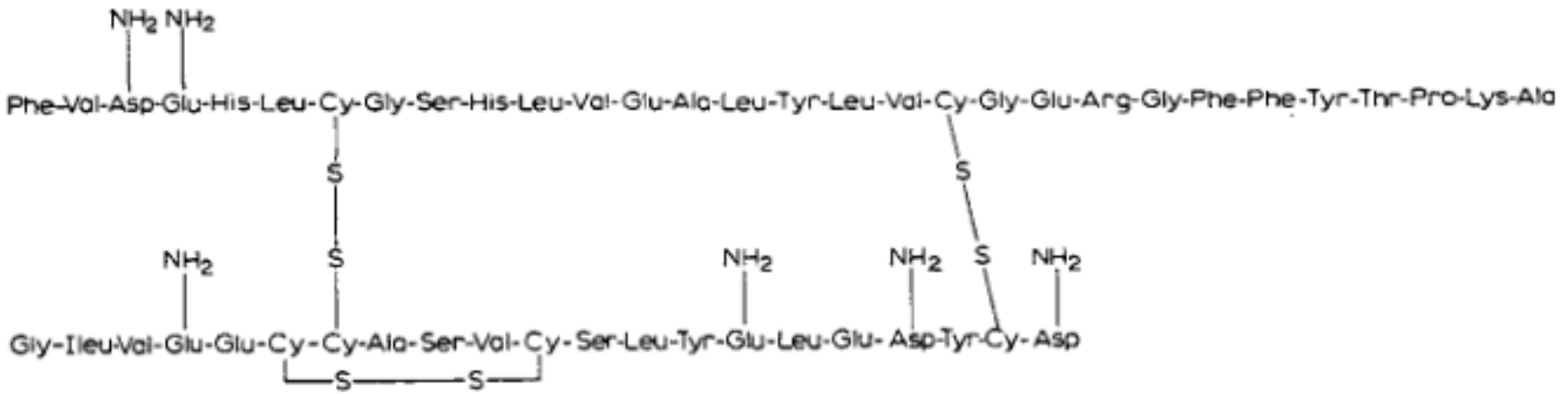


Fig. The structure of insulin.

Of the various theories concerned with protein chemistry our results supported only the classical peptide hypothesis of Hofmeister and Fischer. The fact that all our results could be explained on this theory added further proof, if any were necessary, to its validity. They also showed that proteins are definite chemical substances possessing a unique structure in which each position in the chain is occupied by one and only one amino acid residue.

Examination of the sequences of the two chains reveals no evidence of periodicity of any kind nor does there seem to be any basic principle which determines the arrangement of the residues. They seem to be put together in a random order, but nevertheless a unique and most significant order, since on it must depend the important physiological action of the hormone.

All the above results were obtained on cattle insulin. When insulins from four other species were studied by essentially the same methods it was found that the whole of the B chain was identical in all species and the only differences were found in the three amino acids contained within the disulphide ring of the A chain, which in the cattle are Ala · Ser · Val and in the other species are as follows:

Pig	–	Thr · Ser · Ileu
Sheep	–	Ala · Gly · Val
Horse	–	Thr · Gly · Ileu
Whale	–	Thr · Ser · Ileu

CONTD.

The determination of the structure of insulin clearly opens up the way to similar studies on other proteins and already such studies are going on in a number of laboratories. These studies are aimed at determining the exact chemical structure of the many proteins that go to make up living matter and hence at understanding how these proteins perform their specific functions on which the processes of Life depend. One may also hope that studies on proteins may reveal changes that take place in disease and that our efforts may be of more practical use to humanity.

BIOGRAPHY

Frederick Sanger
The Nobel Prize in Chemistry 1958

Born: 13 August 1918, Rendcombe, United Kingdom

Died: 19 November 2013, Cambridge, United Kingdom

Affiliation at the time of the award: University of Cambridge, Cambridge, United Kingdom

Prize motivation: "for his work on the structure of proteins, especially that of insulin"

Alma mater	St John's College, Cambridge (BA, PhD)
Known for	Determining the amino acid sequence of insulin Sanger sequencing Sanger Centre
Spouse	Margaret Joan Howe[4]
Awards	Nobel Prize in Chemistry (1958) Foreign Associate of the National Academy of Sciences (1967) Royal Medal (1969) Gairdner Foundation International Award (1971) William Bate Hardy Prize (1976) Copley Medal (1977) Louisa Gross Horwitz Prize (1979) Nobel Prize in Chemistry (1980)
Scientific career	
Fields	Biochemistry
Institutions	University of Cambridge Laboratory of Molecular Biology
Thesis	*The metabolism of the amino acid lysine in the animal body* (1943)

WISH U ALL A VERY HAPPY NEW YEAR AND HAPPY PONGAL

-THANK YOU