A new insight into Sanger's development of sequencing

THIS IS AN ADVANCED DRAFT OF A PUBLISHED PAPER. REFERENCES AND QUOTATIONS SHOULD ALWAYS BE MADE TO THE PUBLISHED VERION, WHICH CAN BE FOUND AT:

García-Sancho M. (2010) "A new insight into Sanger's development of sequencing: from proteins to DNA, 1943-77", ***Journal of the History of Biology*****, 43(2): 265-323.**
URL: http://dx.doi.org/10.1007/s10739-009-9184-1

A New Insight into Sanger's Development of Sequencing:
From Proteins to DNA, 1943-1977

Miguel García-Sancho
Department of Science, Technology and Society
Spanish National Research Council (CSIC)
Calle Albasanz, 26-28
28037 – Madrid, Spain[1]
Current email: *miguel.gsancho@ed.ac.uk*

Abstract:
Fred Sanger, the inventor of the first protein, RNA and DNA sequencing methods, has traditionally been seen as a technical scientist, engaged in laboratory bench work and not interested at all in intellectual debates in biology. In his autobiography and commentaries by fellow researchers, he is portrayed as having a trajectory exclusively dependent on technological progress. The scarce historical scholarship on Sanger partially challenges these accounts by highlighting the importance of professional contacts, institutional and disciplinary moves in his career, spanning from 1940 to 1983. This paper will complement such literature by focusing, for the first time, on the transition of Sanger's sequencing strategies from degrading to copying the target molecule, which occurred in the late 60s as he was shifting from protein and RNA to DNA sequencing, shortly after his move from the Department of Biochemistry to the Laboratory of Molecular Biology, both based in Cambridge (UK). Through a reinterpretation of Sanger's papers and retrospective accounts and a pioneering investigation of his laboratory notebooks, I will claim that sequencing shifted from the working procedures of organic chemistry to those of the emergent molecular biology. I will also argue that sequencing deserves a history in its own right as a practice and not as a technique subordinated to the development of molecular biology or genomics. My proposed history of sequencing leads to a reappraisal of current STS debates on bioinformatics, biotechnology and biomedicine.

Keywords: Sanger, sequencing, molecular biology, biochemistry, genomics, recombinant DNA, proteins.

"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 (…) 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 experiments were of paramount importance and fairly narrow specialization was helpful, I managed to hold my own even with the most academically outstanding". [2]

It is an interesting paradox that a leading scientific figure such as Fred Sanger, awarded two Nobel Prizes for his pioneering protein and DNA sequencing methods (1959 and 80), has received little attention by STS scholars. There are, literally, two historical papers on his career [3] and small summaries of his contributions in popular works, mainly on the Human Genome Project. In these latter accounts, Sanger is seen as a precursor of genomics, bioinformatics and biotechnology, by permitting to determine, with his techniques, the sequence of nucleotides in the DNA molecule. Sanger's portrayal as an avant-garde reflects a wider anti-historical view which considers sequencing the culmination of a series of revolutions in molecular biology, starting with the elucidation of the double helix of DNA (1953) and leading, in a straight line, to the recombinant techniques, genomics and the new 1980s biosciences.[4]

This linear and teleological view has been reinforced by Sanger's much quoted autobiographical paper, "Sequences, sequences and sequences," which considers his methods as mainly dependent on the availability, at each historical moment, of instruments for treating the protein, RNA and DNA molecules. In further interviews – including one with myself – Sanger has maintained this interpretation, endorsed in a number of accounts by his collaborators and fellow scientists.[5] The accounts of Sanger and colleagues have informed the popular and growing contemporary STS literature on genomics, which understands sequencing as a series of methods deriving from progress in the recombinant DNA techniques during the 1970s, and allowing, the following decade, the Human Genome Project and other revolutionary endeavors.[6]

The historiography of 20th century life sciences has not yet systematically challenged these accounts.[7] An exception is Soraya de Chadarevian, who in the only historical investigations about Sanger stresses both the complexities in the trajectory of

sequencing – beyond that of recombinant DNA and genomics – and the importance of the research environments surrounding the technologies. Sanger's professional contacts and institutional settings – especially after his 1962 move from the Department of Biochemistry to the Laboratory of Molecular Biology of Cambridge, UK – were essential for the development of sequencing from proteins to RNA and DNA. Equally, the introduction of sequencing into a molecular biology laboratory shaped the emergence of this discipline in Cambridge, the first British city where it was coined.[8]

De Chadarevian, thus, thoroughly analyzes the emergence of Sanger's first protein sequencing methods and their incorporation into the technical repertoire of molecular biology. There is, however, little attention to Sanger's further work on RNA and DNA sequencing during the 1960s and 70s. This period, especially the latter decade, was marked by the emergence of a "new approach" in his sequencing strategy: instead of successively cleaving and then reconstructing the molecule target of sequencing, Sanger attempted to copy it and use the resulting duplicate in deducing the sequence. In her latest paper, de Chadarevian argues that Sanger's early protein techniques "informed initial attempts at nucleic acid sequencing". Nevertheless, little is said about how RNA and DNA sequencing evolved, and especially what was behind Sanger's change of approach.[9]

My paper will address Sanger's transition from degrading to copying procedures and extend the historical investigations on sequencing to RNA and DNA. Earlier scholarship has subordinated the development of sequencing to broader disciplinary moves. De Chadarevian, for instance, uses protein sequencing as a case study to investigate the emergence of molecular biology in Cambridge and the role of biochemists such as Sanger in this process. This paper, in contrast, will inquire when sequencing as a practice emerged and how it developed reciprocally affecting and being affected by disciplinary and professional identities.[10]

When sequencing is understood as a practice, the activities of its inventors and users are stressed and placed in a wider context as particular forms of knowledge and work. It is then easier to link sequencing to researchers who, in solving practical problems, crossed disciplinary boundaries. Sequencing as a technique, on the contrary, has a

less active and dynamic dimensions, being understood as a tool originated and developing within a discipline. My historical approach will, therefore, inquire when scientists started sequencing molecules rather than when sequencing, as a technique, revolutionized molecular biology.[11]

The history of sequencing presented here will require revisiting Sanger's career in light of his scientific papers and, crucially, his laboratory notebooks, recently donated to the Biochemical Society, catalogued at the Wellcome Trust Archives in London and not previously explored by scholars. Both notebooks and papers will be compared with contemporary research lines and placed within the historiography of the "molecularization" of the life sciences.[12] My paper will, consequently, contribute to a better understanding of how biology became molecular – by adopting and transforming a practice such as sequencing. It will also critically assess retrospective accounts of involved researchers, one of the main goals of professional history of science.[13]

The first part of the paper will show that Sanger's early work between the 1940s and 50s was framed in biochemical analysis of proteins. Sequencing at this point lacked any concern with genetics and, as a practice, originated before the elucidation of the double helix of DNA and the emergence of molecular biology. It was not until his 1962 move to the Laboratory of Molecular Biology – addressed in the second part of the paper – that Sanger began directing his techniques to RNA and DNA, and considering their relation to proteins as constituents of the genetic material. This shift, I will argue, did not create a revolutionary *technique* of molecular biology, but rather redefined a pre-existing *practice* according to the parameters of the new emergent discipline.

FIGURE 1

-1.Research on proteins: a new practice emerging in biochemistry (1943-62).

Sanger started his career in a period of profound change at the Department of Biochemistry of Cambridge. Its charismatic director, Frederick G. Hopkins, retired and was substituted by Charles Chibnall in 1943. The broad approach to biochemistry

defended by Hopkins – which included genetic, comparative and embryological studies of metabolism – was substituted by Chibnall's more specific interests in protein analysis. Sanger's early work on insulin initially squared with Chibnall's interests, but gradually developed into a distinctive practice named sequence determination, and directed to the identity and order of amino acids in the protein.

Part 1 of this paper will show that Sanger's early work remained within the confines of biochemistry. The determination of the sequence of insulin was framed in the dominant research line at his department, using the methods of organic chemistry to analyze the amino acid composition of proteins. With the completion of insulin and spread of sequence determination, Sanger's practice began impacting not only Chibnall's team and biochemistry, but also other groups beyond the boundaries of this discipline.

-1.1.Chibnall, Fischer and the structural studies of proteins.

The Dunn Institute of Biochemistry, the name of Cambridge's department at Sanger's arrival, was known for having been created in 1924 by Hopkins, one of the precursors of this discipline. Hopkins defended a "dynamic" or "general" approach to biochemistry, consisting in a multi-perspective study of metabolism (i.e. how nutrients and other inputs were processed by the cell. The research included comparison between the metabolisms of different species, genetic investigations of the regulation of the metabolic process and study of its development from embryo to adult. They were conducted by researchers as diverse as Ernest Baldwin, J.B.S. Haldane or Joseph Needham.[14]

When Sanger started his PhD at the Institute (1940), Hopkins was advanced in years, many of his collaborators had left and his influence was significantly decreased. He was replaced in 1943 by Chibnall, plant biochemist who brought to Cambridge most of his team from Imperial College, London.[15] Historian of biochemistry, Robert Kohler, has shown how Chibnall's arrival substituted Hopkin's broad approach with "much narrower interests in plant proteins". Chemical investigation of proteins became subsequently a privileged research line at the Institute, especially amino acid analysis, a technique of which Chibnall is considered one of the pioneers.[16]

Chibnall described this technique in a 1942 lecture. The proteins were broken into their constituent amino acids by different procedures and then the quantity of each amino acid was determined. Using this method, he had been able to determine the number of "aspartic acid, glutamic acid, arginine, histidine and lysine" in a number of proteins. These studies, Chibnall claimed, would allow the correlation of the "composition" of the protein with its "nutritive value" – i.e. how proteins acted in the body.[17]

Chibnall placed his work within a line of research which had arisen in the mid 19th century and consisted in applying the methods of organic chemistry to the study of biological molecules. In the same lecture, he quoted Justus von Liebig, Franz Hofmeister and Emil Fischer as the sources which inspired his investigations. Historians of biochemistry have considered these German researchers among the pioneers in extending to biology the analytical techniques of chemistry.[18] Hofmeister and Fischer, additionally, are also credited as the postulators of early hypotheses about protein structure.

Both researchers, based on the chemical analysis of a large number of proteins, concluded in 1902 that there were recurrent patterns in their structure. Whereas Hofmeister established that proteins were characterized by repetitive associations of carboxyl and amide groups – i.e. chemical links recurring along their structure – Fischer postulated that they were "chains" of amino acids (*ketie*, in German) linked by "peptide bonds" between such groups. Amino acids, consequently, were associated in small peptide chains.[19]

Historian Joseph Fruton has shown how up to the late 1930s, there were reasons to question Fischer and Hofmeister's hypotheses. First, Fischer himself had expressed doubts about the validity of his peptide bonds in longer protein chains. Second, other researchers considered that for explaining "the specific biological and physical properties" of proteins other types of chemical bond were necessary giving the molecule a globular structure.[20] There was, finally, an enduringly influential research line derived from the mid 19th century, colloidal chemistry, which claimed that proteins lacked a specific molecular structure: they were rather complex mixtures of

materials suspended in fluids – e.g. blood – which did not obey the laws of solution chemistry.[21]

During the 1940s, the debate was being gradually solved in favor of the molecular structure of proteins. Chibnall, the research of other protein chemists, and the emergent field of physical chemistry had proven that even in long protein chains, the linear arrangement of discrete amino acids was the most plausible structure. Nevertheless, even amongst the supporters of Fischer and Hofmeister there were disagreements on the nature of the amino acid arrangements in the peptide chains characterizing proteins.

-1.2.Periodical or undetermined chains?

Fischer's peptide hypothesis had different interpretations among its followers during the 1930s and 40s. A considerably widespread one at the time of Chibnall's appointment in Cambridge was that of Max Bergmann, former collaborator of Fischer working at the Rockefeller Institute of New York. Bergmann was Head of the Institute's Laboratory of Chemistry where he investigated protein analysis and synthesis. Taking a series of observations by his assistant Carl Niemann, Bergmann postulated in the late 30s the periodicity hypothesis, according to which protein chains were formed by patterns of amino acids which were repeated at regular intervals.[22]

Bergmann further proposed the formula 2^n X 3^m, which allegedly described the total number of amino acids in any given protein (Bergmann and Niemann, 1938, pp. 577-579). The powers (*n* and *m*) of 2 and 3 defined the different intervals at which the amino acids were, supposedly, repeated. Taking the equation, Bergman and Niemann proposed "a classification of the numerous individual proteins" in which their basic properties and structure would be mathematically estimated through variations of the formula 2^n X 3^m.[23]

Not all researchers, however, interpreted Fischer's theory in the same way. Among the most critical biochemists of the periodicity hypothesis were those based in Cambridge who were convinced that proteins were undetermined structures. Their

study, hence, required chemical analysis instead of mathematical predictions. Bill Pirie and Albert Neuberger, Sanger's PhD Supervisors, were among the first in attacking Bergmann and Niemann during the late 1930s, also using mathematical arguments. In independent papers, they showed that almost every possible distribution and total number of amino acids in a protein may be expressed as a result of the formula 2^n X 3^m.[24]

Chibnall also dismissed the periodicity hypothesis, but used chemical analysis instead of mathematical methods. In his 1942 paper, he showed that the amino acid composition of edestin, β-lactaglobulin, egg albumin and insulin yielded values contradicting Bergmann and Niemann's formula. Based on this, he concluded that the protein molecule was a "system of peptide chains of varied composition" and that, for determining its structure, it was necessary to analyze it physically and chemically, rather than mathematically:

> "Clearly the analyst (…) can contribute but little on his own, and there is need for more co-operation with the physical chemist and crystallographers, who are able to investigate the properties and structure of (…) the intact protein molecule itself. (…) Meanwhile, I think that those interested in proteins would be wise to regard the Bergmann-Niemann hypothesis as still tentative and in any case as applicable only to the component peptide chains of the molecule, for much of the evidence hitherto brought forward to support it has been based on inadequate experimental data and has demonstrated nothing more than the hypnotic power of numerology".[25]

This context of critique to periodicity and preference for chemical methods framed Sanger's early work on insulin during the mid and late 1940s. Departing from Chibnall's amino acid analysis, he substantially modified the technique and this allowed him not only to determine the "composition" of proteins, but also the "sequence" of their amino acids along the chain.[26]

-1.3.Determining an unpredictable sequence.

Chibnall's arrival in Cambridge coincided with the completion of Sanger's Ph.D. thesis. The similarities in their personalities, both preferring the bench to the meeting room, resulted in trust and affinity. Chibnall asked Sanger to participate in his project on chemical analysis of insulin, started at Imperial College and which became the main line of research in his group at Cambridge. Sanger accepted the offer and was

incorporated to the team, but instead of just applying the amino acid analysis techniques, he introduced a series of variants which determined not only the number and nature, but also the position of the amino acids in the molecule.[27]

Sanger's first assigned duty within the insulin project was quantitatively estimating and identifying the amino acids located at the end of the molecule's chains. For so doing, he devised a method combining procedures already used by other group members with new instruments. Among the former, he applied column chromatography, a technique which permitted the separation of the amino acids of a protein after their being cleaved from the chain. This method had been introduced in the early 1940s and extensively applied to amino acid analysis by Chibnall's team, among other researchers.[28]

Sanger, however, added to the method a chemical substance – dinitrofluorobenzene (DNFB) – suggested by Chibnall and not used before in protein analysis. DNFB reacted with the last amino acids of the insulin chains, dying them yellow. This substance allowed Sanger not only to quantify, but also to identify the amino acids, in line with Chibnall's request. By applying DNFB and cleaving the insulin molecule with acids, he broke all the amino acids of the chains. He then separated the mixture of loose amino acids through column chromatography and determined, from their yellow color, which ones were located at the end of the chains.[29]

The paper in which Sanger reported the results of those experiments, written in 1945, defined insulin as a molecule formed by "four open polypeptide chains".[30] He was, then, describing the protein as a series of peptide-linked amino acid chains, in line with Chibnall's investigations and Fischer's hypothesis.[31] In his subsequent papers, published in 1949, Sanger began referring to the insulin molecule as a "sequence,"[32] a term which despite having been used before (see next subsection) had never appeared in Fischer or Chibnall's articles. Sanger's sequences were formed by the ends of the insulin chains together with the neighboring amino acids, to which he directed his further experiments.

Sanger concretely managed to identify and determine the position of the four to five amino acids beside the ends of the insulin chains between 1945 and 49. He achieved

this by introducing two innovations into his technique. Firstly, the protein was submitted to partial rather than complete cleavage: instead of separating the amino acids individually, Sanger reduced the time of acid exposure and obtained peptides, i.e. mini-chains of four to five amino acids. Secondly, an improved separation method called paper chromatography was incorporated, which involved a sheet of cellulose as the medium and performed the separation in two instead of one direction. The resulting amino acid pattern on the paper – i.e. the chromatogram – was, hence, two-dimensional.[33]

The new method, which Sanger named the degradation approach, consisted in submitting various insulin samples to partial cleavage and the resulting yellow peptides – DNFB labeled and at the end of the chains – to further breakages. In this way, he obtained a series of overlapping amino acid fragments, for instance phenylalanine, phenylalanine-valine and phenylalanine-valine-aspartic acid. By separating the fragments through chromatography and submitting them to amino acid analysis – as Chibnall and his group members were doing – it was possible not only to identify and quantify, but also to deduce the order of amino acids in the chains. "The increasing complexity of each of those peptides," Sanger claimed, "suggested that they were all breakdown products of the same peptide chain".[34]

FIGURE 2

Sanger referred to these amino acid peptides as *sequences*. The term differed from his 1945 paper, and the use of the term *chain* by Fischer, Chibnall and other members of his group. Sanger, nevertheless, used his sequences as evidence for the protein structure hypothesis preferred by his Institute. In the conclusions of his 1949 papers, he mentioned that the investigations on insulin and comparison with the terminal peptides of other proteins[35] suggested that there were no rules governing their structure and it was, consequently, unpredictable:

> "Investigations of the free amino groups of a number of proteins by [this] technique (...) have shown that the terminal position in the protein chains may be occupied by a variety of different amino acids. There appears to be no principle that defines the nature of the residue occupying this position in different proteins and it would seem probable that this would apply to other positions in the molecule".[36]

Sanger's early experiments on insulin were, therefore, framed in the perspectives and research interests of Chibnall's group, but also progressively shifted towards new goals. On the one hand, Sanger exclusively used chemical methods to analyze the protein chains and suggested with his experiments that they were unpredictable, in line with Cambridge's Dunn Institute views.[37] On the other, in so doing he gradually abandoned his first duty of determining the terminal amino acids of insulin and directed his methods to increasingly larger *sequences* within the molecule. Sanger's latter goals resulted in a new practice within biochemistry, identified during the 1950s with sequence determination.

-1.4.An emerging practice.

By introducing further modifications into the method, Sanger managed to apply it to the central parts of insulin and to determine its whole sequence between 1951 and 55.[38] At that time, the use of the term *sequence* increased in Sanger's papers. Whereas in previous articles this term had been either absent or combined with other ways of referring to the protein chains such as "free amino groups," "terminal peptides" or insulin "structure", in these new investigations the papers clearly stated in their titles that the object of Sanger's research was the "amino acid sequence" of the molecule. Furthermore, Sanger described his activity as "to determine the complete amino acid sequence" of certain parts of insulin – firstly its phenylalanyl and then its glycil chain.[39] He also began to benefit from the help of his assistants, biochemists Hans Tuppy and E.O. Thompson, who co-authored the papers, learnt the sequence determination techniques and spread them in a variety of ways during their further careers.

FIGURE 3

Sanger was not the first in using the term sequence for referring to protein structure. Archer Martin and Richard Synge, also biochemists educated in Cambridge, had raised the possibility of studying "structural sequences" in proteins during the mid 1940s. The parallelisms between these researchers and Sanger are remarkable. Synge had also been based in the Dunn Institute of Biochemistry and supervised by Pirie just before Sanger's arrival – mid-late 1930. Martin had started his career at an offshoot

of that institute, the Dunn Nutritional Laboratory. Both investigators, as Sanger, were devoting their research to the development of methods, in this case for the separation of amino acids and other biological mixtures.[40]

The career of Sanger has significant overlaps with those of Martin and Synge, based from the early 1940s in the Wool Industries Research Institution in Leeds.[41] There, Martin and Synge introduced column and paper chromatography, the main techniques for the experiments Sanger reported in his 1945 and 49 papers. Sanger, additionally, coincided with Synge in 1947 during a visit to Arne Tiselius's laboratory in Uppsala, which was another major center in the development of chromatography and separation techniques.[42]

Martin and Synge's investigations, like those of Sanger, had similarities with the work of Chibnall's group during the early 1940s. When they first proposed determining structural sequences of proteins in 1943, their separation techniques were directed to the "composition" of the small peptide gramicidin. The paper sought to chemically identify and estimate quantitatively the amino acids forming gramicidin, and the determination of its sequence was only stated as a long-term possibility. However, with the introduction of paper chromatography in 1944, Martin and Synge presented this new separation technique as directed to "qualitative analysis of proteins". The technique's capacities, they suggested, went beyond a simple quantification of amino acids. In 1947, Martin and Synge used paper chromatography to determine the "sequence" of gramicidin and proposed applying the technique "with increasing confidence for studies of the sequences of amino acids residues in peptide structures generally".[43]

Sanger had also been working on gramicidin during the 1940s, at the same time as his early insulin work. Despite only publishing one paper on gramicidin, the first volumes of his notebooks – written between 1945 and 49 – contain a number of experiments on this peptide scattered among those on insulin. Furthermore, prior to his 1949 papers – when Sanger first used the term *sequence* – he began referring to the results of paper chromatography in the notebooks as "qualitative." All these parallelisms – together with the extensive quotes to Martin and Synge in Sanger's

1950s papers – suggest that Sanger adopted the concept of sequence from these two researchers.[44]

There were, however, differences between Sanger, and Martin and Synge's sequence determination efforts. Whereas the latter used sequence determination as a means to test their separation techniques, Sanger made this activity the goal of his career. Sanger's research objects were the sequences; the methods to determine these included not only separation techniques, but also other biochemical instruments. He was the first in making sequence determination a goal and devoting the whole efforts of his research to it.

Sequence determination, therefore, emerged as a new practice in Britain during the mid and late 1940s. Martin and Synge defined it as a distinguishable activity, and Sanger transformed it into a clear career objective. This career objective overcame the quantitative amino acid analyses dominant at Cambridge's Dunn Institute of Biochemistry and directed them towards a new goal through a differentiated methodology.[45] Nevertheless, despite these peculiarities, Sanger's insulin method remained strongly framed in the research tradition represented by Fischer and embodied in Chibnall's group. Sequence determination, initially, incorporated the working procedures of analytical and synthetic organic chemistry.

-1.5.A purely (bio)chemical activity.

Bergmann, Chibnall, Martin and Synge, and Sanger, despite their discrepancies, shared a working tradition marked by the application of the instruments of chemistry to biological molecules. This tradition had been established, among others, by Fischer, a chemist by background who between the late 19th and early 20th century became increasingly interested in proteins and applied to them his techniques. The application of chemistry to biology and proteins was progressively named biochemistry – or more concretely protein chemistry – being the research environment in which Sanger developed his work.

In a 1907 lecture, Fischer reflected on the convergence of chemistry and biology. He considered it natural and quoted classical chemists – among them von Liebig – as the

precursors of this confluence.[46] At the beginning of the 20th century, Fischer claimed, chemistry was equipped with a "powerful armory of analytical and synthetical weapons" which would bring a "clearer insight" into the "processes which constitute animal and vegetable life". These methods, developed in their modern version between the late 18th and 19th centuries, were respectively characterized by breaking molecules into fragments – analytical – and assembling such fragments into new compounds – synthetic.[47]

Fischer had pioneered the application of these techniques to proteins, analyzing and synthesizing a large number of compounds. He named his analytical techniques "degradation methods" and directed them to the "composition" of the molecules – just as Sanger and Chibnall would do.[48] In his 1907 lecture, Fischer described his experiments as follows:

> "The conclusions which have been drawn in other cases from the results obtained by the dissection of compounds have been too frequently confirmed by their synthesis. It is now possible to make this claim on behalf of the proteins, as it has been found to be possible, by a process the reverse of hydrolysis [breakage of the chains] to associate amino acids in such a manner that substances are produced which (…) resemble proteins. I have termed these synthetic products polypeptides".[49]

Fischer's analytical and synthetic tradition was followed by Bergmann and Chibnall. The former, after working as a close associate of Fischer, created a research school at the Rockefeller Institute, engaged since the 1930s in the synthesis and structural analysis of proteins. Chibnall, contemporaneously, formed a group devoted to the study of the amino acid composition of insulin and other proteins, moving it to Cambridge in the early 1940s.[50]

In his history of biochemistry, Kohler has identified "three distinct styles" characterizing the diversity of this discipline during the first decades of the 20th century: clinical, biological and bioorganic-biophysical.[51] Fischer's line of research and its continuation by Bergmann and Chibnall corresponded with the third one which was strongly oriented towards analytical and synthetic chemistry. None of these researchers showed medical concerns or any interest in the biology (e.g. the metabolic processes) of living organisms.

Chibnall's 1943 move to Cambridge may, hence, be interpreted as a shift from a biological to a bioorganic and biophysical style at the Dunn Institute of Biochemistry. This researcher lacked the broad approach to metabolism of Hopkins and as new Head of the Institute favored the application of the analytical techniques of chemistry to proteins.[52] His model was, however, short-lived, since in 1949 Chibnall resigned due to the large administrative duties involved in the position and the Institute turned to a more medical orientation.[53]

Sanger's sequence determination efforts were decisively shaped by the growth of protein chemistry at the Institute and remained an important line of research after Chibnall's tenure.[54] The degradation procedure was based on the breakage of insulin into fragments and its further reconstruction through identifying overlaps. In so doing, Sanger was applying the cleaving, separation and quantification methods of analytical chemistry already used by the other members of Chibnall's group. He was also employing, though not experimentally, the strategy of synthetic chemistry in attempting to reconstruct the original molecule from the fragments' overlaps.

Sequence determination, as a distinctive practice, also added specific features to Cambridge's protein chemistry. By incorporating new instruments to amino acid analysis – DNFB and paper chromatography – Sanger gave a new qualitative dimension to this method, able to yield since then the *sequence* and not only the *composition* of insulin. He also created, with Martin and Synge, the concept of sequence and the aim of achieving it through the development of techniques. Both the concept of sequence and the aim of determining it became especially pervasive during the 1950s and increasingly spread among the biochemical community.

-1.6.Spread and convergence with molecular biology.

The progressive completion of insulin (1951-55) and the award of the first Nobel Prize to Sanger (1959) made sequence determination expand in protein research. During the 1950s and 60s, an increasing number of biochemists either adopted the technique or developed alternative sequence determination methods. In both cases, researchers equally assimilated the working principles Sanger's technique embodied:

their use of sequence determination was based on the instruments and strategic approaches of analytical and synthetic chemistry.

One of the first investigators in devising sequence determination techniques after Sanger was Pehr Edman, a Swedish biochemist based in the University of Lund. His method, developed in 1950, also started with the cleavage of the protein, but instead of acids or enzymes, Edman employed a chemical substance able to cut the last amino acid of the chain. By successively applying this substance and submitting the cleaved amino acid to paper chromatography, Edman's technique could determine sequences unit per unit in the correct order. This procedure was, consequently, also inspired by the "degradation" of the protein and the term was recurrently used in Edman's paper.[55]

Edman's technique – rather than Sanger's – was adopted as the basis for the automation of sequence determination. The process started in the late 1950s, with the Rockefeller Institute researchers William Stein and Stanford Moore designing an apparatus which automated amino acid analysis, i.e. the identification and quantification of the protein breakage products after being separated through chromatography. Stein and Moore's strategy, described as "subtractive," consisted in successively cutting the terminal amino acids of the protein by Edman's technique and checking, after each cleavage, which unit was missing through the automatic analyzer. It was, again, based on the degradation of the molecule and the activity which was automated – amino acid analysis – was exactly the same Chibnall's group had been performing in its investigations of protein composition.[56]

At the same time Stein and Moore were automating amino acid analysis, Sanger was seeking techniques for avoiding this step in his protein methods. Amino acid analysis required applying a series of reagents to the chromatogram, in order to make the protein units visible after separation. The spots were then identified and quantified. Sanger considered this process "time consuming and tedious," and after the determination of insulin (1955) sought means to skip it and to deduce the sequence directly from the chromatogram or alternative separation surface.[57]

His main strategy, developed between the late 1950s and early 60s, was to label the protein fragments and amino acids with radioactive substances. After separation, Sanger photographed the chromatogram or alternative medium, so that the radioactively labeled fragments or amino acids appeared as dark bands. The sequence could, in principle, be deduced by measuring the position of the spots in the picture, called a protein fingerprint.[58]

FIGURE 4

Radioactive labeling had been introduced to Sanger by Chris Anfinsen, a United States biochemist who had been visiting the department in 1954.[59] The use of this technique dates back to the 30s in the fields of physiology, nuclear medicine and various biological disciplines. During the 50s, radioactive labeling was applied in the Hershey-Chase experiment and that of Matthew Meselson and Franklin Stahl, which combining biochemistry, bacteriology and genetics respectively showed that DNA is the genetic material as well as the replication mechanism of this molecule.[60] Sanger was, thus, using a technique bridging a variety of biological disciplines – including the then emerging molecular biology – for eliminating amino acid analysis, an important chemical step in sequence determination.

Another research line opened by Sanger in the late 1950s aimed to deduce the functioning of insulin from its recently determined sequence. He did it by comparing sequences of different species – e.g. pig and sheep – and studying previously radioactively labeled active centers of enzymes – i.e. proteins involved in the catalysis of chemical reactions.[61] The interest in comparing and the postulation of a linear relationship between molecular structure and function had been present since the foundation of the Dunn Institute and reinforced by Chibnall's protein analysis program. Historians of biology Sage Ross and Bruno Strasser have shown – in still ongoing investigations – that this straight view of the structure-function connection was characteristic of comparative biochemistry, a research tradition present at the Institute since Hopkins's time, in the late 1920s and 30s. Baldwin, its main representative, had been Sanger's undergraduate tutor and an important influence on his further career in biochemistry.[62]

The hypothesis of a straight connection between molecular structure and function was also emerging, throughout the 1950s, in biophysics. This discipline, articulated around the technique of x-ray crystallography (see notes 7, 20 and 29), had gained momentum with the elucidation of the double helix of DNA (1953) which led some researchers to focus their experiments on this molecule with the expectation of deducing the mechanisms of gene functioning. Philosopher of biology Sahotra Sarkar has characterized this hypothesis as reductionism, showing that it shaped the history of the emergent discipline of molecular biology. During the late 50s and 60s, an increasing number of biophysicists declared themselves the founders of molecular biology and built their further investigations on this DNA-centered scheme.[63]

Cambridge was also an important center for these converted biophysicists. Since the late 1950s, Sanger maintained increasingly intense contacts with a group of crystallographers studying biomolecular structures at the Cambridge-based Cavendish Laboratory. None of Sanger's post-insulin research lines had led to visible results at that time.[64] The apparent stagnation of Sanger's protein work, together with the growing contacts with the Cavendish researchers, led him to start a new stage in his career by moving, in 1962, from the Dunn Institute to a new molecular biology center and applying his sequence determination techniques to RNA and DNA.

Sanger's move triggered a major reconfiguration of sequence determination as a practice. It consolidated the shift witnessed since the late 1950s and redefined his techniques in terms of the emergent molecular biology. This redefinition will be analyzed in the following section, which will further challenge accounts explaining Sanger's career exclusively in terms of technological progress and the success of molecular biology. The investigation so far conducted has initially challenged such arguments by showing that Sanger's institutional and professional environment was also crucial in providing 1) *a line of research on the chemical analysis and synthesis of proteins* epitomized by Fischer and developed by Chibnall's group; 2) *already available analytical methods for amino acid analysis* combined by Sanger with other instruments in the development of his techniques; 3) *an aim of determining the sequence of insulin* which led Sanger to innovatively combine the available instruments and to the emergence of sequence determination as a practice initially framed in biochemistry.

-2.Research on nucleic acids: from chemical degradation to biological replication (1962-77).

The impact of Sanger's determination of insulin and his subsequent Nobel Prize (1955 and 59, respectively) made researchers all around the world follow his work. Especially active in this was the Cavendish Laboratory, a center in Cambridge which since the late 1940s had shifted from physics to host an expanding group of researchers studying the "molecular structure of biological systems," as well as other related problems.[65] Francis Crick, Sydney Brenner and other Cavendish scientists saw Sanger's techniques as especially suitable for their investigations on protein synthesis or, as they also called it, the problem of the genetic code: how DNA specifies the structure of proteins. They approached Sanger in the late 1950s and persuaded him to move to a new center, the Laboratory of Molecular Biology (LMB), which was being specifically built to combine the Cavendish biological group with other researchers investigating related problems. After the move, Sanger began applying his techniques to RNA (1960s) and then to DNA (1970s).

In all his retrospective accounts, Sanger has acknowledged the role of the Cavendish biologists in his move to the LMB. However, he has argued that these contacts "did not achieve the experimental level" and that his career remained mainly shaped by the "availability" of sequence determination instruments.[66] Part 2 of this paper will challenge this interpretation and show how the new research environment of the LMB affected Sanger's techniques, especially in their transition from RNA to DNA. Since the 1960s, I will argue, sequence determination gradually shifted from the instruments, problems and strategic approach of chemistry to those of the emerging discipline of molecular biology. This transition, however, was not one-directional, as autobiographical and some historical accounts of molecular biology suggest. In its journey, sequence determination as a practice also had an impact on the way of conducting molecular and other forms of biology.[67]

-2.1.The move to a new center.

Sanger's interactions with the biological group at the Cavendish Laboratory – named Unit for the Study of Molecular Structure of Biological Systems – have been studied in detail by Soraya de Chadarevian and reviewed in autobiographies by the involved researchers.[68] These interactions started in the late 1940s and were especially intense from 1955 onwards. At that time, Sanger had already gained scientific prestige, was introducing radioactive labeling into his techniques and attempting to determine the properties of insulin from its sequence. Both radioactive labeling and the hypothesis of there being a connection between the structure and function of biological molecules were also important in biophysics and molecular biology, the nascent field that the Cavendish group was promoting at that time.[69]

The Cavendish scientists were mainly physicists who had shifted to biological problems after World War II. They were applying to biological molecules – namely proteins – the technique of x-ray crystallography, in order to determine their three-dimensional structure. The strong orientation of this institution towards physics had prevented Sanger – especially antipathetic towards this discipline (see note 32) – to establish consistent links with it, despite his research interests being apparently close. Nevertheless, the elucidation of the double helix of DNA (1953) and the increasing attraction of biochemists towards this molecule transformed the situation in the mid-late 50s.[70]

A key figure in this transformation was Francis Crick. Physicist and crystallographer by background, he had cooperated with James Watson – a bright young US researcher visiting the Cavendish – in the elucidation of the double helix, reported in two epoch-making papers in 1953. At that time, Crick was already aware of Sanger's techniques and used the successively published results on insulin as evidence for the formulation of the sequence hypothesis. The hypothesis stated that given the complexity of protein structure – as Sanger had demonstrated in insulin – only an equivalently intricate molecule as DNA could generate it. This led Crick to postulate that the sequence of bases in DNA determined the sequence of amino acids in proteins.[71]

FIGURE 5

Despite the sequence hypothesis not being published until 1958 in a paper devoted to "protein synthesis", it inspired Crick's work on the genetic code from the mid 1950s. In collaboration with another Cavendish researcher, Sydney Brenner, he investigated how the DNA molecule, in its expression, determined a particular sequence for proteins. During the late 50s, Crick and Brenner produced a series of mutations in simple organisms – bacteriophage viruses – and studied how their altered sequences of DNA affected the formation of proteins.[72]

De Chadarevian has shown how Brenner and Crick soon realized that Sanger's techniques would be extremely useful for matching concrete mutations in the DNA of the viruses with changes in the amino acid sequence of the resulting proteins. This led to contact and then an informal cooperation between Sanger and Brenner, in which the former determined the sequence of the proteins produced by the mutant viruses obtained by Brenner. A more formalized and productive association was that of Sanger with Vernon Ingram, a member of Crick's group investigating sickle-cell anemia. Ingram used Sanger's late protein techniques to identify the alteration in the amino acid sequence of hemoglobin causing the disease in 1956.[73]

The potential of these collaborations made Crick and Brenner continually attempt to attract Sanger to the Cavendish Laboratory from the mid 1950s. They wanted him to keep applying his protein techniques to the genetic code and to begin developing methods for sequence determination of RNA and DNA. The efforts of both researchers intensified when at the end of the decade the Medical Research Council, a body of the British Government funding biomedical sciences, decided to establish a new center, the LMB, in which the Cavendish group would share building with other biologists, being Sanger an essential incorporation.[74] The biological nature of this institution overcame Sanger's past reticence towards physics and made him accept the offer, moving to the LMB in 1962, when it opened.

Before the move, Brenner and Crick organized two seminar sessions in the Golden Helix – residence of the latter – where they taught the fundamentals of molecular biology to Sanger and the members of his group, also due to migrate. The lectures included the basic structure of DNA, its relation to RNA and proteins in the genetic

code, and, crucially, the mechanism of cell replication. Sanger and Crick's recollection of this period differs in their retrospective accounts: whereas the former seems to have forgotten the seminars and claims to have acquired all his background on nucleic acids from another LMB member, John Smith, Crick considers the Golden Helix lectures as a key element in Sanger's move.[75] This apparently unimportant divergence reflects Sanger and Crick's different view of the emergence of molecular biology.

-2.2.Molecular biology and Sanger's professional identity.

Sanger's move to a molecular biology center and introduction into this emerging field did not prevent him from considering his career as fully integrated in biochemistry. In his retrospective accounts, he has understood his research as invariably "involved in the same project" – sequence determination – which survived institutional moves and other environmental changes. When commenting on the effects of the move, Sanger refers to the "larger space and facilities" of the LMB, without seeing significant changes in his work. The LMB, according to Sanger, was named Laboratory of Molecular Biology because "there was already a department of biochemistry in Cambridge." Sanger's lack of distinction between both disciplines is confirmed when he refers to Crick as a "theoretical biochemist".[76]

Sanger's account contrasts with de Chadarevian's study of the foundation of the LMB. For de Chadarevian, Sanger's move was a key episode in the emergence of molecular biology in Cambridge, marked by the "alliance of protein crystallographers, molecular geneticists and protein chemists." This convergence required "institutional and disciplinary negotiations," as shown by Sanger's collaboration with Crick, Ingram and Brenner, together with the opening of the LMB, specifically built to gather these scientists[77]. De Chadarevian, hence, argues that the history of Sanger at the LMB is not one of a biochemist preserving his identity within molecular biology, but of a biochemist contributing, with researchers from other fields, towards the emergence of a new discipline.

However, in the accounts of that emergence, Crick and Brenner's views have prevailed over those of Sanger. The Cavendish researchers have propagated the idea

that Sanger invented the techniques, but they saw "the problems to which those techniques could be usefully applied". This interpretation, indirectly supported by Sanger's retrospective accounts, squares with the dominant picture of Sanger as a pure technologist applying his instruments to the theories formulated by leading molecular biologists.[78] It is also a unidirectional account, in which the effects of the interactions between Crick, Brenner and Sanger are seen as having operated in only one way, exporting a series of inert techniques to a dynamic new field. Crick and Brenner's personalities – eloquent and charismatic, in contrast with Sanger's shyness – have reinforced the dominance of their account.[79]

This points towards a more general problem in the historiography of 20th century life sciences: the pervasiveness of the retrospective accounts of molecular biologists. The prestige and perceived revolutionary nature of this discipline has led to the proliferation of autobiographies in which prominent molecular biologists offer their interpretation of the development of the field. These accounts, influential among the lay public and in contemporary STS research alike, have shaped the perception of biology after 1953 as dominated by the "revolutions" around the DNA molecule.[80] Historians have partially mitigated this effect by placing molecular biology within the context of the development of post-World War II life sciences. However, as de Chadarevian and Harmke Kamminga have argued, some scholarship on "molecularization" has identified the emergence of a molecular vision of life with the rise of molecular biology between the 1950s and 70s. This literature raises a unidirectional model of interaction in which all the biological disciplines were transformed according to the parameters of molecular biology.[81]

Sanger's early career shows that the phenomenon of molecularization cannot be solely explained in the development of molecular biology. Whilst still within the working parameters of biochemistry in the 1950s, he incorporated instruments and strategies which approached sequence determination to molecular biology. Radioactive labeling and the belief in a straight structure-function relationship bridged sequence determination as a practice and the interests of the Cavendish biophysicists, later self-declared molecular biologists. Furthermore, Ingram's use of fingerprinting techniques showed Sanger the connection between protein sequence alterations and a genetic disease such as sickle-cell anemia. Another key factor in

Sanger's introduction into the realm of molecular biology was his interest in protein synthesis since the late 1950s.

-2.3.Protein synthesis and the transition to nucleic acids.

An analysis of Sanger's laboratory notebooks shows that his transition from protein to nucleic acid sequence determination was gradual and not just fostered by a single event. His latest insulin volumes, written between 1960 and 62, show that his experiments on RNA started while still at the Dunn Institute, in parallel with those on radioactive labeling and comparison of protein sequences. At that time, Sanger became interested in the problem of protein synthesis and incorporation of amino acids into the protein by transfer RNA. In his RNA books, written after his move to the LMB, the experiments on proteins decreased, but the engagement with synthesis remained to inform Sanger's work even during the 70s, when he shifted to DNA sequence determination. Some of his LMB assistants – e.g. Ieuan Harris and Brian Hartley – kept working on proteins in the 60s and 70s.[82]

FIGURE 6

Protein synthesis had been a concern of biochemists since the first decades of the 20th century, including researchers at the Dunn Institute especially before Chibnall's arrival. In the 1950s, it also became a main interest of molecular biologists, for it epitomized the relationship between DNA, RNA and proteins. Historian Bruno Strasser has shown how during the late 40s and 50s there were alternative models of protein synthesis proposed by both biochemists and molecular biologists. Hans-Jörg Rheinberger has argued that during this same period, protein synthesis as an "experimental system" gradually shifted from biochemistry to be used by an increasing number of self-declared molecular biologists[83]. Protein synthesis was a concern of both Sanger and Crick between the late 50s and early 60s. However, whereas the latter postulated it as a one-directional transfer of information from DNA, RNA and then proteins – the celebrated central dogma of molecular biology – [84] Sanger regarded it rather as a means to link his protein work with that which he had begun on nucleic acids.

Sanger's engagement in protein synthesis since the late 50s shows that his contacts with Crick and Brenner, as well as Ingram's experiments, were important in interesting him in the concerns of molecular biology. However, the previous tradition on protein synthesis in biochemistry, together with the way Sanger introduced it in his notebooks – as well as radioactive labeling and sequence comparison – reflects a gradual incorporation rather than a dramatic shift. Sanger assumed protein synthesis for practical reasons, to give coherence to his work and not as a disciplinary dogma. This suggests that he was not converted to molecular biology in a single non-critical step – e.g. the Golden Helix seminars or his move to the LMB. Sanger rather progressively incorporated tools and strategies of molecular biology to facilitate his transition to nucleic acids.

Sanger's move to RNA and DNA sequence determination, hence, endorses de Chadarevian and Kamminga's model of molecularization: two-directional and relying on particular research strategies. Both historians have argued that the molecularization of the life sciences between the 1950s and 60s depended not only on molecular biology, but on previous research traditions which were equally centered in a molecular explanation of biological phenomena.[85] In this regard, Sanger's contacts and move to the LMB transformed sequence determination from its previous biochemical identity, but sequence determination as a practice also impacted the way in which molecular biology was conducted. Sanger's techniques, consequently, had a meaning of their own and transformed the problems to which Brenner and Crick applied them. In order to develop this argument, I will first review the changes Sanger introduced into his techniques while at the LMB, in order to explore their historical meaning.

-2.4.The development of a new technique.

Sanger's first task at the LMB was expanding his already initiated experiments on RNA sequence determination. He did so with a similar method to that he used in his late insulin work. The RNA fragments were degraded with enzymes and radioactively labeled, so that the sequence could be deduced from the separation surface – also two-dimensional – by detecting overlaps.[86] Sequence determination, hence, remained mainly within the parameters of analytical and synthetic chemistry.

With this technique, Sanger determined the sequence of a number of short transfer and ribosomal RNAs during the 1960s and early 70s. Sanger's investigations attracted increasing interest and contacts from molecular biologists, not only at Crick and Brenner's laboratory, but also outside the LMB. During the second half of the 1960s and 70s, a number of molecular biology institutions devoted considerable efforts to determine longer RNA sequences.[87] Sanger's assistants at that time, equally, had less specific biochemical profiles: Bart Barrell, joining the laboratory in the early 1960s, was educated to A-level without a university background, and George Brownlee, appointed at the same time, came from a biological degree in Cambrdige.

When Sanger attempted to apply the degradation procedure to DNA, he found that it was unsuitable to this molecule, due to DNA being longer than proteins and RNA. This forced him to create an "entirely new approach" to sequence determination, called the *copying procedure* and based on duplicating rather than degrading the molecule. In this method, developed during the late 1960s and 70s, the DNA was no longer successively broken, but replicated with an enzyme (polymerase) which, in the presence of a series of loose nucleotides, progressively added complementary bases to the DNA to be determined – i.e. the template DNA.[88]

By different means – the so-called plus and minus, and dideoxy methods – Sanger managed to stop the polymerase additions at each of the four nucleotides. Since polymerase acted simultaneously in various DNA samples, the result was a series of overlapping fragments in which the last unit (adenine, cytosine, guanine or thymine) was known. If the fragments were separated by size, the sequence could be deduced by analyzing the ends of the fragments – progressively bigger – on the resulting separation surface.[89]

FIGURE 7A

The copying procedure incorporated gel electrophoresis, a different separation technique from chromatography. By placing the DNA fragments on a slippery surface – a porous gel – and applying an electric charge, gel electrophoresis was able to displace them according to size with remarkable precision (Thurtle, 1998). The

separation gel electrophoresis produced in DNA was different from chromatography. Sanger divided the gel into four lanes and placed the fragments respectively ending in adenine, cytosine, guanine and thymine in each of them. When the electric charge was applied, the fragments migrated vertically from top to bottom. The separation, hence, was one-dimensional instead of two-dimensional, with the fragments ending in each nucleotide vertically aligned and horizontally separated into four lanes.[90]

As a result of the fragments being radioactively labeled and having a ladder-like overlapping nature, the sequence could be deduced by scanning a picture of the gel – called autoradiograph – with the eye. The researcher began at the bottom of the picture, checking for a dark band – the product of radioactivity – in the four lanes. By progressively moving up and repeating the operation, it was possible to determine the sequence.

FIGURE 7B

-2.5.The insufficiency of a technological explanation.

Sanger has persistently claimed that his shift from a degradation to a copying procedure was mainly due to the different "availability of technologies" between the 1960s and 70s. Both in his autobiographical account and further interviews, he has stressed the role of a series of technical instruments which, allegedly, became available at the time he developed his DNA methods. These instruments, namely polymerase and gel electrophoresis, were, according to Sanger, the crucial factors in his change of approach once he realized that the degradation procedure was unsuitable for DNA.[91]

Sanger's account is, to a large extent, based on his self-definition as a "technologist," i.e. a scientist concerned with the "development of methods" rather than with the formulation of hypotheses.[92] However, a closer look at his transition to copying suggests that this account may have been created to fit with Sanger's self-portrayal. The main evidence for this is Sanger's hesitation about this matter in a passage of his autobiographical paper. In it, he admits that the account may have been

retrospectively constructed, since he does not have sound enough memories of the time:

> "[The] new approach to DNA sequencing was I think the best idea I have ever had (…), 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 [my assistants] Alan Coulson and John Donelson. I have a feeling that the above account [entirely technological] 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".[93]

Sanger's account may, thus, be a simplification in light of the further success of the copying procedure. It presupposes a Eureka moment – having the copying *idea* – and a single shift in which the method was developed by adopting the available technologies. The technological explanation, therefore, squares with Crick and Brenner's view of Sanger entering into molecular biology and changing his sequence determination procedures in one simple step. Nevertheless, the late 1960s and 70s record of experiments shows that they were conducted in a radically different fashion.

Sanger's account, firstly, considers only his successful experiments. However, his laboratory notebooks show that Sanger's failed attempts and dead ends were at least as significant as his viable procedures which did not develop in a linear fashion.[94] The experiments are frequently accompanied by comments such as "where everything goes?,", "don't seem to make much sense" or "look ghostly", referring to the chromatograms or autoradiographs.[95] These comments and parallel alternative attempts became especially abundant around 1966, when Sanger and his assistants first directed their techniques towards DNA.

The team first aimed to create hybrid templates of RNA and DNA, adopting polymerase since, at least, 1970. Its intention then was not to pursue a copying procedure, but to create shorter DNA fragments in order to apply the degradation strategy. These polymerase-derived fragments were submitted to two and then one-dimensional electrophoresis by John Donelson, a US biochemist who pioneered the application of gel electrophoresis at the LMB during a postdoctoral stay. Nevertheless, his goal was to reconstruct the sequence from the overlaps rather than

to scan it visually, as in the copying approach. Sanger and his assistants published a number of papers with small DNA sequences elucidated through these techniques.[96]

Around 1973, Sanger reported in his notebooks an attempt consisting in "copying with all four" DNA nucleotides. Shortly afterwards, he described a "fairly ambitious experiment" which incorporated the main features of the copying procedure. The method required a significant number of refinements in subsequent experiments and the expressions "plus" and "minus approach" gradually gained significance. Also gradually, the experiment reports acquired a homogenous format and a model DNA sequence – that of bacteriophage virus ØX-174 –was determined using the new copying approach.[97]

FIGURE 8

Sanger's copying approach, therefore, emerged gradually rather than as a product of a Eureka moment. As with his transition from protein to nucleic acids, it was a consequence of practical decisions while developing his sequence determination techniques and not an abrupt shift towards molecular biology. Sanger's career, in this regard, presents the features of what F.L. Holmes has called "investigative pathway," a research trajectory which does not develop in a linear fashion, but shows an overall continuity.[98] What was important in Sanger's investigative pathway was not the technologies or their availability, but rather the reasons for his combining them in the copying approach.

This is especially true since the same technologies were being simultaneously used during the 1960s in sequence determination and other structural analyses of nucleic acids. Researchers in the United States and Europe were applying polymerase and gels to RNA and DNA at the same time Sanger was developing his copying approach. This contemporary use, apart from further questioning the availability of technologies argument, shows that sequence determination was becoming a widespread practice, with multiple ramifications and approaches. Its study, thus, requires focusing on all the historical attempts rather than on retrospectively big names, such as Sanger.[99]

Polymerase had been isolated in 1956 and applied to RNA and DNA sequence determination in the subsequent decade.[100] Molecular biologists at Stanford and Cornell University A.D. Kaiser and Ray Wu used it in the late 1960s to determine the sequence of the sticky ends of the bacteriophage λ, two short single-stranded DNA fragments at the ends of its genome. One year later, the Zurich-based researcher Carl Weissmann applied a variation of the enzyme to the RNA virus Qβ. This enzyme, RNA polymerase, was able to duplicate ribonucleic acids.[101]

Kaiser, Wu and Weismann were, thus, creating an approach to sequence determination based on reproducing a biological process – RNA and DNA replication with polymerase – rather than degrading and reconstructing the molecule. This approach was, however, combined with other non-biological instruments to deduce the sequences: Kaiser and Wu used statistical analyses to measure the incorporation rates of the different nucleotides by polymerase, whereas Weissmain applied the two-dimensional chemical separation.[102] In the case of Sanger, the transition towards biology would be more significant.

A key element for Sanger's transition was the adoption at the LMB of *gel electrophoresis* as the separation method. This instrument had been introduced by Tiselius in the 1930s and extensively applied to proteins.[103] During the 60s, a number of researchers, largely with medical motivations, used gel electrophoresis for qualitative and quantitative analyses of nucleic acids. They did so by dividing the gel into lanes and running a different RNA or DNA molecule – previously radioactively labeled – in each lane. By comparing their different mobility, they estimated the "species" of nucleic acid molecule in the lanes.[104]

Sanger had sporadically used electrophoresis in different media – especially paper – during the development of his protein and RNA techniques. However, it was not applied systematically in his laboratory until the first attempts to determine DNA sequences in the late 1960s. Donelson – who had been a student of Wu – was the first to apply electrophoresis at the LMB, but mainly from within the degradation approach. With the advent of the copying procedure, Sanger introduced a key modification in the handling of the gel: instead of running various DNA molecules, one in each lane, he produced, from a single DNA template, fragments ending in each

of the four nucleotides and used the lanes to separate the nucleotides rather than the molecules. This way, the lanes were not representative of a species of DNA, but of a group of fragments ending, respectively, in adenine, cytosine, guanine and thymine.[105]

FIGURE 9

Gel electrophoresis and polymerase were, thus, differently applied to sequence determination by Sanger rather than introduced for the first time into this practice. Both instruments had a recognizable tradition of use in sequence determination and other biological areas when Sanger incorporated them to his DNA copying procedure. This suggests that, rather than their availability, the most significant factor was their being combined innovatively at the LMB in a new approach to sequence determination.

This approach, additionally, started to be referred to as *sequencing* by Sanger and the increasing number of researchers using this practice at the LMB and abroad. The term first emerged in Sanger's early 1970s laboratory notebooks and was used in the paper where he described the plus and minus method, published in 1975. It was then incorporated into the title of the article "DNA sequencing with chain terminating inhibitors," where Sanger described the dideoxy technique (1977).[106] The term spread among molecular biologists during the 80s and became the dominant way of describing sequence determination. Sequencing, as well as being called by a different name, had significantly changed its identity when compared with Sanger's previous degradation approach.

-2.6.Sequencing and its move towards biology.

The insufficiency of a technological explanation raises the question of what made Sanger combine polymerase and gel electrophoresis in the way that resulted in his DNA copying approach. A tentative answer is the novel research environment of the LMB, where Sanger was immersed since 1962 after starting introducing into his work the problems of DNA and its relation to RNA and proteins. If the dominance of chemical analysis and synthesis at the Dunn Institute had shaped Sanger's

degradation procedure, the new research problems, methodologies and instruments used at the LMB may have created the context which, beyond the technologies, affected the emergence of copying.

This context has been analyzed in detail by de Chadarevian, who defines the LMB as a "federative structure" marked by three independent groups. The first two had been operative at the Cavendish Laboratory and respectively focused on x-ray crystallography and the emergent discipline of molecular genetics. Whereas the crystallography group – leaded by Perutz and Kendrew – had traditionally focused on proteins and maintained this interest at the LMB, the molecular geneticists – headed by Crick and Brenner – increasingly focused on DNA and its connection to RNA and proteins. The third group, led by Sanger, entered the laboratory in 1962 and simultaneously developed protein and nucleic acid sequencing techniques. [107] Nevertheless, Sanger increasingly focused on the latter and left proteins to his assistants.

During the early and mid 1960s, coinciding with the foundation of the LMB and Sanger's move, the molecular genetics group was shifting towards new research areas after the elucidation of the genetic code mechanism in 1961.[108] One of these was cell division, pursued by Brenner in cooperation with François Jacob, another emergent molecular biologist at the Pasteur Institute in Paris.[109] The entrance of Brenner and Jacob into this problem, according to the former, led to a shift in the investigation of cell division from the perspective of biochemistry to that of molecular biology.

Brenner supports this claim by pointing to a change of the methods they used to investigate how DNA directs the copy of the cell by replicating itself. Shortly after the isolation of polymerase in the mid 1950s, a number of biochemists had put this enzyme and DNA into a test tube, with the aim of analyzing the process. Brenner and Jacob, rather than artificially reproducing the event, sought to understand how it worked in a living organism which they had already used, the bacterium *E. coli.*[110] In their investigations, which began in 1962, they also applied radioactive labeling and a strategy Brenner had used in his 1950s genetic code research: producing mutations in *E. coli's* DNA and analyzing their effects in the replication and cell division processes.[111]

As these experiments were being conducted (1962-63), Sanger was developing his RNA technique through a procedure which was still framed in chemical degradation. His work was closely followed by Brenner, who maintained regular exchanges with Sanger and used RNA sequence determination in his cell division experiments. Brenner characterized *E. coli* mutants by seeking alterations in the sequences of the messenger RNA produced by the mutated genes.

The members of Sanger's group during the 1960s kept shifting from pure biochemistry to a more general biological specialization. Alan Coulson arrived to the laboratory in 1967 after finishing a degree on Applied Biology in Cambridge. Coulson, with Donelson, was within the group of assistants more directly involved in the early development of the DNA techniques. The fact that his appointment was more permanent made Coulson stay in the laboratory until the emergence of both, the plus and minus and dideoxy method, and sign the papers as co-author (1975 and 77).

Sanger's laboratory notebooks show that Coulson's role in the "fairly ambitious" experiments which triggered the copying approach was decisive. Coulson was involved in the first attempt of using polymerase with the four DNA nucleotides and its further refinement between 1973 and 75. Furthermore, the emergence of the plus and minus method in the notebooks coincides with Donelson's departure from the group after his attempt of using polymerase and gels within the degradation procedure. When reflecting retrospectively on the copying approach, Coulson considers that its success and pervasiveness was mainly due to the approach reproducing "the natural functioning of the cell".[112]

Coulson's statement suggests that, during the first half of the 1970s, sequence determination shifted from its previous biochemical identity and began being modeled on the process of DNA replication. This problem had been a main focus of the neighboring molecular genetics group and Brenner, its main representative, had been cooperating with Sanger throughout the preceding decade. In devising the copying approach, Sanger and Coulson adopted the instruments and experimental strategies of molecular biology, dominant at the LMB and embodied in the *natural* division of the cell.

Sanger's immersion in molecular biology, though, was gradual and started with his move to the LMB. Shortly after this, he began using as the main publishing medium of his papers the *Journal of Molecular Biology* instead of the *Biochemical Journal,* where he had published most of his protein work. The *Journal of Molecular Biology* had been founded by another LMB member, John Kendrew, becoming the reference publication for researchers in the new field. There, Sanger published his RNA and DNA techniques, and the references to problems of molecular biology – such as protein synthesis or cell division – increased throughout the 1960s and 70s. A 1975 lecture delivered shortly after the publication of his plus and minus method started with the following statement:

> "DNA, the chemical component of the gene, plays a central role in biology and contains the whole information for the development of an organism, coded in the form of sequences of the four nucleotide residues.[113] The lecture describes the development and application of some methods that can be employed to deduce sequences in these very large molecules".[114]

Sanger has argued that he assumed this language and problems for them being the "gospel" of molecular biology. According to this, he just referred to the concepts and concerns of his new home institution, without their affecting his working procedures. In his experiments, Sanger claims, "it did not really matter" how he was "thinking on DNA." The really important issue was the development of sequencing methods through the adoption of the available technologies.[115]

It is probably true that Sanger was unaware of his transition to molecular biology. However, an analysis of his copying procedure shows that it decisively affected the development of sequencing as a practice. Firstly, in both the plus and minus and dideoxy methods, Sanger used part of the DNA nucleotides and the polymerase of *E. coli,* the same model organism as Brenner and Jacob.[116] Secondly, instead of chemically intervening in the DNA template by breaking it with acids or enzymes, he just left polymerase to act on the molecule and to reproduce the natural process of DNA replication. Thirdly, as Brenner and Jacob had done with the *E. coli* mutations, Sanger directed the replication process towards his desired outcome by selectively stopping the action of polymerase on the DNA template. Fourthly, he introduced radioactive labeling and electrophoresis for the purpose of visualizing the sequences,

without having to chemically analyze them. And, finally, around 1975 Sanger started applying his methods to the DNA of a living organism, the virus ØX-174, of the same type (bacteriophage) of those Brenner and Crick had used in their late 1950s genetic code experiments.[117]

Sanger was, hence, incorporating into the practice of sequencing the methodology, instruments and problems that molecular biologists had used in the study of cell division. The development of his DNA copying approach was triggered by a transition from the working procedures of chemistry to those of molecular biology, embodied in the molecular genetics group of the LMB, his new home institution.[118] While during the 1960s Sanger had kept within the parameters of analytical and synthetic chemistry, the contact with Brenner's research questions and role of assistants such as Coulson led him to introduce replication as the basis of DNA sequencing the following decade. This made DNA sequencing different from the previous techniques, modeled on the opposite process of chemical degradation.

Sanger's sequencing methods, however, still maintained some of its original chemical features. The DNA template, firstly, was not replicated *in vivo,* but *in vitro* and the development of the dideoxy method – in which the nucleotides were modified to stop the action of polymerase (see note 47) – required a large amount of analytical and synthetic chemistry. Secondly, unless the DNA was unusually short, the molecule needed to be cleaved previously with enzymes in overlapping fragments. These fragments were then sequenced independently through the copying procedure and, finally, reconstructed. Sanger, thus, kept part of the original biochemical identity of sequencing within the copying approach.

Other researchers developing alternative sequencing methods during the 1970s combined biochemistry and molecular biology differently. Nevertheless, Sanger's combination better satisfied the necessities of molecular biologists, who felt more comfortable with processes involving replication through polymerase than degradation using chemicals. This partially accounts for the pervasiveness of Sanger's techniques in the 80s and their dominance over rival methods.

-2.7.Rival attempts and the idea of elegance.

At the same time Sanger was developing his plus and minus and dideoxy methods, an alternative approach to DNA sequencing was being invented in the United States by Walter Gilbert with the help of Allan Maxam. Gilbert was a molecular biologist based in Harvard, where he had been working with Watson on the genetic code. Nevertheless, he had a strong interest in chemistry and the development of his sequencing techniques was inspired by the Soviet biochemist Andrei Mirzabekov, who was visiting Gilbert's laboratory in 1975.[119]

This meant that Gilbert's technique combined chemistry and biology in different proportions than at the LMB. Gilbert, as Sanger, used radioactive labeling and gel electrophoresis to produce a band pattern which did not need to be chemically analyzed. However, instead of applying polymerase, he submitted the DNA sample to chemicals which specifically cleaved the molecule at each nucleotide. The idea of using such chemicals was specifically suggested by Mirzabekov during his 1975 visit. By successively cleaving four groups of identical DNA samples at adenine, cytosine, guanine and thymine, Gilbert obtained the same overlapping fragments which Sanger was achieving with polymerase. The rest of the process was the same as that of the LMB: the fragments were separated through electrophoresis and the sequence deduced from the band pattern formed on the autoradiograph.[120]

Gilbert's technique, despite its being available since 1975, was published in the same journal issue as Sanger's dideoxy sequencing in 77. Both methods were tested in a number of laboratories during the late 1970s and 80s. The assessment looked initially favorable to Gilbert, since Sanger's technique could only be applied to single-stranded DNA, while the Harvard method had no such limitation. However, the further automation of DNA sequencing in the mid 1980s was based on Sanger's procedure, which relegated Gilbert's technique to a secondary role.[121]

Neither researchers nor scholars have been able to explain this preference on technical grounds alone. Whereas Sanger talks about being "lucky," Coulson refers to his boss's method as "more elegant".[122] By elegant, he means that the procedure was more easily applicable to a laboratory during the late 70s and 80s. The fact that the

LMB was a molecular biology laboratory implied that the idea of elegance referred to the capacity of Sanger's sequencing to adapt to the working procedures of this discipline, wherein the bulk of its early users were found.

This leads, again, to Coulson's remark that Sanger's sequencing reproduced the natural functioning of the cell. Researchers in molecular biology were used to these sort of procedures, which incorporated the natural workings of DNA into their experiments. Brenner had used with Crick and then with Jacob the mechanisms of DNA expression and replication in his attempts to solve respectively the genetic code and cell division during the late 1950s and 60s (see Sections 2.1 and 2.6). This meant that, for these investigators, a technique such as Sanger's, which used polymerase as a duplication agent seemed more natural and elegant than a method which degraded DNA with chemicals. On the contrary, Gilbert's technique was perceived as compromising the security and working standards of a laboratory. Molecular biologists, unsurprisingly, referred to it as "the chemical method" of sequencing during the late 1970s and 80s.[123]

Sanger's sequencing was, thus, preferred for its broader engagement with the procedures of molecular biology, which were inspired by the *natural* workings of DNA. This led to Gilbert's technique being progressively abandoned and to its not being used in the large-scale sequencing projects arising at the end of the 1980s. Sanger's dideoxy method, by contrast, became the standard sequencing procedure, was further automated and had an increasing impact on key areas of biological research.[124]

The spread of Sanger's technique made molecular and other biological researchers direct their investigations increasingly towards the sequence of information in the DNA of different organisms and use information technologies – such as the computer and the database – for that endeavor. The large-scale sequencing initiatives arising in the late 1980s – among them the Human Genome Project – were a reflection of this new goal, given unprecedentedly generous funding by political and scientific institutions (ibid., pp. 133-48). Sanger's sequencing, in retrospect, consequently looks like a technique arising from the 1970s recombinant DNA technologies and paving

the way to genomics, the new mapping and sequencing field. However, the history offered in this paper makes the development of sequencing look less straightforward.

-4.Conclusion: sequencing and the new histories of genomics.

This paper has challenged the standard accounts of Fred Sanger's career in order to portray protein, RNA and DNA sequencing as distinctive practices, suitable to historical analysis. Its main argument, building on Soraya de Chadarevian's scholarship and challenging Sanger's autobiography as well as other popular accounts, is that the development of sequencing (1940 to 83) cannot be solely explained in the technologies Sanger had available at a given time. The research environments of the Dunn Institute of Biochemistry of Cambridge (UK) and the Laboratory of Molecular Biology (LMB) played a crucial role in Sanger's transition from a degradation to a copying approach to sequencing shortly after his 1962 move to the latter institution. Such transition can be seen as a gradual shift from the working procedures of chemistry – cleavage and reconstruction of molecules – to those of molecular biology – use of the natural replication mechanisms of DNA. Sequencing, conceived in this way, emerges as a distinctive and gradually developing practice framed within and changing across particular research environments.

Understanding Sanger's sequencing as a practice rather than as a technique – or techniques – deriving from the formal development of disciplines has implications for the historiography of molecular biology and genomics.[125] It shows that the history of sequencing is longer and more complex than that of molecular biology and the revolutions around DNA. Sequence determination, in this regard, emerged within protein biochemistry and did not interact with molecular biology until the late 1950s, more than a decade after Sanger's initial experiments on insulin. It also spread among biochemists, molecular and other biological researchers, becoming a generalized practice developed not only by Sanger, but by his expanding assistants and independent researchers which included Pehr Edman, A.D. Kaiser or Ray Wu.

This long and neglected history of sequencing, firstly, is a suitable case study for de Chadarevian and Harmke Kamminga's model of "molecularization" of 20th century

life sciences (de Chadarevian and Kamminga, eds., 1998). It shows that in the decades following World War II, there was a two-directional interaction between molecular biology and other pre-existing biological practices, something that the traditional historiography on molecularization has overlooked (Abir-Am, 1982; Kay, 1993; Olby, 1990). Sequencing was, consequently, shaped by the problems, instruments and experimental strategies operating at the LMB – as leading molecular biologists have claimed – but also had an impact on the way of conducting molecular and other forms of biology since the 1960s.

Secondly, by looking at sequencing as a historicized practice, genomics emerges as something more than a revolutionary consequence of the recombinant DNA techniques. If the aims, strategies and orientation of genomics – large-scale mapping and sequencing – are seen from the perspective of evolving practices, they are no longer exclusive to the late 1980s, but rather the result of the gradual development of sequencing and other biological practices which predated DNA and molecular biology. The history of genomics, therefore, should be sought in the distinctive histories of such practices and not only in the celebrated molecular *revolutions* or the technological progress following the 1970s recombinant techniques.

My paper is, thus, framed within an incipient historiographical trend looking at the diversity of 20th century biology and, more concretely, aiming to build a historical perspective around genomics (Gaudillière, 2002; Rheinberger and Gaudillière, 2004ab; Suárez and Ramillon, eds., upcoming). Historians are starting to show "lineages" and "continuities" between the practices of genomics and those from evolutionary biology, natural history, comparative biochemistry or the post-Fordist organizational models of the factory (Suarez, 2007, 2008; Strasser, 2006a, 2008; Pickstone, 2007; Ramillon, 2007; Bonneuil and Gaudillière, 2007). Their accounts offer an alternative to the autobiographies of molecular biologists, particularly pervasive and which are reproduced in popular literature as a straight line of progress from 1953 – elucidation of the double helix of DNA – to genomics and the Human Genome Project (e.g. Watson, 2003; Gilbert, 1992; Cook-Deegan, 1994; Judson, 1992).

This historical scholarship also expands the ongoing contemporary STS research on genomics. A major concern within this field is the so-called ELSA – studies on the Ethical, Legal and Social Aspects of genomics – not just following genomics research, but also making an impact on its agenda (Goven, 2008; Tutton, 2008). For this purpose, it is essential that current debates in the politics, economics, sociology and philosophy of science – e.g. gene patenting, eugenics or medical translation of research results – overcome the discourse constructed by the public and genomics researchers (e.g. Hood, 1992; Gilbert, 1992). My paper, and the historiographical current within which it is framed, show the potential of history to meet such an aim. Only by looking at the distinctive histories of the practices constituting genomics is it possible to see the whole scope of the construction of this field and its associated debates. The long history of genomics this paper has proposed is, consequently, essential for seriously challenging the myth of the revolutionary nature of this field.[126]

Acknowledgements

David Edgerton and Max Stadler (Centre for the History of Science, Imperial College London); Soraya de Chadarevian (Center for Society and Genetics, UCLA); Adam Bostanci and other researchers at the ESRC Centre for Genomics in Society (University of Exeter, UK); Catherine Heeney (ETHOX Centre, University of Oxford); José Manuel Sánchez Ron, Javier Ordóñez, Antonio Sillero and Rafael Garesse (Universidad Autónoma de Madrid); John Pickstone and other researchers at the Center for the History of Science (University of Manchester); María Jesús Santesmases (Department of Science and Society, CSIC, Spain); Jerôme Pierrel (University Louis Pasteur, Strasbourg); Richard Ashcroft (Queen Mary University, London); Hans-Jörg Rheinberger, Edna Suárez and Vincent Ramillon (Max Planck Institute for the History of Science, Berlin); Joe Cain and Catherine Jackson (Science and Technology Studies Department, University College London); Bruno Strasser (Yale University); Manfred Laubichler (Arizona State University); Annette Faux (Laboratory of Molecular Biology of Cambridge Archives); Fred Sanger, Bart Barrell, Alan Coulson (retired, Sanger Institute and Laboratory of Molecular Biology of Cambridge).

I would like to especially thank my PhD Supervisor at Imperial College, Andrew Mendelsohn, as well as three anonymous referees for key suggestions to refine my arguments.

John Lagnado, Honorary Archivist of the UK Biochemical Society, Jane Kirby and Christopher Burns, Archivists at the Wellcome Trust, also deserve special thanks for granting me access to Fred Sanger's laboratory notebooks before their cataloguing being completed.

The research on which this paper is based was conducted while holding postgraduate fellowships awarded by Caja Madrid Foundation, Madrid City Hall and Residencia de Estudiantes (Spain), as well as a Hans Rausing Fellowship awarded by the Centre for the History of Science, Imperial College, London. Without them, it would not have been feasible. The fieldwork trips and other expenses were also covered by small grants awarded by the Royal Historical Society. The final stages in the preparation of the manuscript were supported by a Wellcome Trust postdoctoral fellowship at the Centre for the History of Science, University of Manchester, and a contract by the Spanish National Research Council (CSIC) at its Center for Humanities and Social Sciences.

Notes

[1] A substantial part of the investigations reported in this paper were conducted during the development of my PhD at the Centre for the History of Science, Imperial College, London, and a short postdoctoral stay at the Centre for the History of Science, University of Manchester.
[2] Sanger, 1988, pp. 1-2.
[3] De Chadarevian, 1996, 1999.
[4] E.g. Judson, 1977, 1992; Wills, 1991; Cook-Deegan, 1994.
[5] Autobiography: Sanger, 1988. Interviews: id., 1987, 1992, 2004; Lagnado, 2005; Hartley, 2004; Coulson, 2004; Stretton, 2002.
[6] E.g. Kevles and Hood, 1992; Sloan, 2000.
[7] Despite problems such as internalism, whiggism or technological determinism having long been addressed in history of science, the standard scholarship on post-World War II biology has not approached this phenomenon historically. On the one hand, the literature on molecular biology focuses on the foundational period of this discipline (1950s and 60s), considering protein sequencing only for its contributions to the achievements around DNA (Abir-Am, 1982, 1992; Olby, 1992 [1974], Fox Keller, 1995; Morange, 1998). On the other hand, histories of previous and contemporary disciplines such as biochemistry address only protein sequencing, with the further RNA and DNA techniques being a matter of molecular biology (Fruton, 1972; Kholer, 1982, Florkin, 1972).
[8] De Chadarevian, 1996, 1999. Other historical works in preparation which also address Sanger are those of Bruno Strasser and Sage Ross, scholar and PhD student at the Department of History, University of Yale. They both partially investigate Sanger's early career to analyze the tradition of comparative biochemistry (Strasser, 2008; Ross, 2008).
[9] Sanger, 1988, pp. 20 and ff.; de Chadarevian, 1999, p. 203.
[10] De Chadarevian, 1996, especially pp. 382 and ff. Further initiatives in the direction of my paper are emerging, especially in the realm of protein sequencing. The workshop "Making Sequences Matter," which took place in June 2008 at the University of Yale, explored the use of early sequencing methods and protein sequences by evolutionary biologists between the 1950s and 70s.
[11] My focus on scientific practice is inspired by John Pickstone's understanding of the history of science, technology and medicine as the development of recurrent, changing and interacting "ways of knowing" and "working" – or "working knowledges" in his latest scholarship: natural history-craft, analysis-rationalization and experimentalism-systematic invention (Pickstone, 2001, especially pp. 7-20; 2007). The development of sequencing as a practice involved all Pickstone's categories. In incorporating them or changing the focus from one to the other, it crossed traditional disciplinary boundaries between the 1940s and 70s.
[12] Kay, 1993; Abir-Am, 1982; Rheinberger, 1998; de Chadarevian and Kamminga, 1998.
[13] Abir-Am and Elliot, 1999; Holmes, 2001, 2006.
[14] Weatherall and Kamminga, 1992; Kohler, 1982, pp. 73-89; Ross, 2008; Paul, 1984; Armon, 2006; Abir-Am, 1982, pp. 361-67.
[15] Synge and Williams, 1990; Fruton, 1992, pp. 35-36.
[16] Kohler, 1982, pp. 88-89; Weatherall and Kamminga, 1992, pp. 80-81; Gay, 2007, pp. 148-149 and 171-172.
[17] Chibnall, 1942, p. 137.
[18] Chibnall, 1942, pp. 136-37; Florkin, 1972, vol. 30, chs. 7 and 15; Fruton, 1972, ch. 2; Kohler, 1982, ch. 2.
[19] Hofmeister, 1902; Fischer, 1902; Fruton, 1985, pp. 315-16, 326-330. Other important German figure at that time was Albrecht Kossel, who postulated "the necessity for the quantitative accounting of a protein" in terms of its amino acids based on Fischer and Hofmeister's hypotheses (Fruton, 1979, p. 5). Chibnall, however, does not quote Kossel in his 1942 lecture on amino acid analysis.
[20] Fruton, 1979, quote p. 10. Many of these alternative structures (e.g. the cyclol hypothesis) were inspired by x-ray crystallography, a parallel research line aimed at the three-dimensional structure of proteins. Chibnall's group, however, did not particularly interact with researchers into this field, which had an important school in Cambridge, with figures such as John Kendrew or Max Perutz (see notes 20 and 29).
[21] Florkin, 1972, vol. 30, pp. 279-85; Creager, 1998, pp. 111-16; Edsall, 1979.
[22] Fruton, 1979, p. 13 and ff.; Clarke, 1944, pp. 169-70. Bergmann and Niemann's hypothesis had been inspired by William Astbury, crystallographer at the University of Leeds and one of the first proponents of periodicity-based theories. In a mid 1930s paper, Astbury claimed that "it is exceedingly

uncommon (...) for molecules to be disorganized in the solid state" and that "proteins (...) are not simply very long, but are also periodic polypeptide chain systems" (Astbury, 1934, pp. 15 and 23).
[23] Bergmann and Niemann, 1938, pp. 582-584, quote 582. Fruton – who in his previous career as a biochemist worked with Bergmann – has claimed that the periodicity hypothesis was "welcomed by some geneticists," since it set "a mathematical principle that might link Mendel's laws to the structure of proteins" (Fruton, 1979, p. 13). The Rockefeller Institute had been a major setting for the early investigations on the nature of the genetic material, believed to be proteins during the 1930s and 40s. Phoebus Levene, former Director of the Institute's Laboratory of Chemistry and well connected with Fischer, proposed the tetranucleotide hypothesis in 1910, which like the periodicity hypothesis for proteins postulated a repetitive parameter in the structure of nucleic acids (Olby, 1992 [1974], pp. 81 and ff.). This, together with the will by the Rockefeller Foundation – owner of the Institute – to apply mathematical physics to biology, suggests that Bergmann and Levene's hypotheses were a product of their time rather than unfortunate *errors,* as some historians have stated (e.g. ibid., pp. 81-82).
[24] Neuberger, 1939, pp. 25-26, quote p. 26; Pirie, 1939, pp. 351-53. Sanger's PhD thesis addressed the metabolism of the amino acid lysine. Despite being unrelated to his later insulin work, Sanger has always claimed it gave him the necessary chemical expertise to develop his techniques (Sanger, 1942, pp. 2 and ff.).
[25] Chibnall, 1942, quotes from pp. 158 and 159.
[26] Quotes from Chibnall, 1942, p. 137 and Sanger, 1949ab. The contrast between Bergmann and Niemann's mathematical approach to proteins and that of the Dunn Institute squares with a similar dichotomy investigated by Lily Kay in the context of the genetic code. According to her, research on protein synthesis during the 1950s and early 60s shifted from cryptographic methods to those of biochemistry, exactly as I have shown it happened with structural analysis of proteins twenty years before (Kay, 2000, chs. 3-6).
[27] Chibnall, quoted in Fruton, 1992, pp. 35-36. Sanger's early methods have been widely described both autobiographically and by other scholars, but with a limited focus on their technical features (Sanger, 1959; 1988, pp. 5-11; Wills, 1991, pp. 29-33; Stretton, 2002; Fruton, 1999, pp. 215-18). The account I will offer, in line with the overall orientation of this paper, seeks to frame Sanger's technical choices in his institutional and professional environment. Ross has also described Sanger's early methods in some detail with the aim of characterizing a discipline – comparative biochemistry (Ross, 2008).
[28] Sanger, 1988, pp. 4-5; Morris, 1998; Gordon, 1977; Martin, 1952, p. 366.
[29] Chibnall, quoted in Fruton, 1992, p. 36; de Chadarevian, 1999, p. 203; Sanger, 1945, pp. 507-515; 1988, pp. 4-6.
[30] The number of insulin chains was a main research question at the time and the aim of Chibnall's initial request to Sanger (de Chadarevian, 1999, p. 203; Chibnall, quoted in Fruton, 1992, p. 36). In 1945, it was thought that insulin was composed by four instead of two chains. This was due to an estimation of its molecular weight which was higher than our present standards (Sanger, 1988, pp. 6-7; 1949b, pp. 157-59).
[31] Sanger also describes Fischer as a fundamental source of inspiration in all his retrospective accounts (Sanger, 1988, pp. 1-2; 2005). His laboratory notebooks are accompanied by a volume of Fischer's collective papers, given to Sanger by the German scientist's son – H.O.L. Fischer – at a 1951 conference in Berkeley. The volume holds the following dedication: "To Dr. Fred Sanger in commemoration to his impressive presentation of his wonderful results in the Chemistry of Insulin" (Sanger's Laboratory Notebooks, Wellcome Trust, London, item number SA/BIO/P/5).
[32] Quotes from Sanger, 1945, p. 514 and id., 1949a, p. 563.
[33] Sanger, 1949a, pp. 564-65; 1949b, pp. 154-57.
[34] Sanger, 1949a, p. 556; 1949b, pp. 154-57, quote in 155; Wills, 1991, p. 30.
[35] Whilst focusing his investigations on insulin, Sanger simultaneously worked on other proteins, such as globin, edestin, hemoglobin and gramicidin (see next subsection). These investigations, which did not generally result in publications, have experiments devoted to them in his notebooks, and even a specific volume in the case of globin (Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Books 5 to 13; Book on Globin, reference number SA/BIO/P/1/1).
[36] Sanger, 1949a, p. 573. Sanger has stressed in further accounts the importance of his 1949 paper for the "philosophy of proteins" (Sanger, 1987a; 1992). Nevertheless, he has retrospectively forgotten having suggested the idea of an unpredictable structure at that time (Sanger and Dowding, 1996, p. 7). This, together with the lack of STS investigations on the above conclusion, is symptomatic of the narrowly technical approach that scholars have adopted when analyzing Sanger's career.

[37] The influence of the Dunn Institute's view of protein structure in Sanger's work has also been suggested by Fruton. Nevertheless, in 1952, Sanger still considered the unpredictable nature of proteins – and even the peptide theory – hypothetically (Fruton, 1992, pp. 43-44; 1999, pp. 215-18; 1972, p. 1).
[38] Sanger, concretely, began combining acids with enzymes in the cutting of the chains. The enzymes had the advantage of breaking the molecule at specific sites –e.g. between valine and alanine – and permitted, upon identification of the breakage point, knowledge of the structure at the beginning and end of each fragment –valine and alanine. This allowed the application of the degradation procedure to the central parts and not only to the edges of the protein (Sanger, 1951; 1953b, pp. 353-374; 1959, pp. 1340-44; 1988, pp. 9-11).
[39] Quotes from Sanger, 1945; 1949ab; 1951ab; 1953ab and 1953b, p. 23. The transition towards the term "sequence" is, nevertheless, slower in Sanger's laboratory notebooks. Whereas the term appears combined with concepts such as "peptides" or "chains" in the last volumes devoted to proteins, from 1950 onwards, "sequence" becomes more prominent in those devoted to RNA and DNA, written after his move to the Laboratory of Molecular Biology in 1962 (Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Books 7 and ff.; RNA and DNA Books, files number SA/BIO/P/1/20 and ff.). This suggests that Sanger rethought his terminology when presenting his experiments in papers and, during the 1950s, decided to name his work homogeneously as sequence determination. The terminological gap between notebooks and papers is also visible in his RNA and DNA work (see below).
[40] Quote from Gordon, Martin and Synge, 1943; see also Weatherall and Kamminga, 1992, pp. 64-66 and 83. The parallelisms between Sanger, Martin and Synge suggest a particular orientation in Cambridge's biological centers, engaged at that time in the development of techniques to determine molecular structures. Cambridge's engagement is also reflected in the influential school of x-ray crystallography developed there since the 1930s, with figures such as J.D. Bernal or Lawrence Bragg. James Watson and Francis Crick deduced the double helical structure of DNA partially from crystallographic pictures in Cambridge in 1953 (Olby, 1992 [1974], Part IV). Sanger, however, was not very interested in crystallography (see notes 7 and 29).
[41] Bergmann had also been working in the chemical industry, concretely at the Kaiser Wilhelm Institute for Leather Research in Dresden during the 1920s, before moving to the Rockefeller Institute (Clarke, 1944). Sanger, in contrast, did not have any connection with industry at any time in his career, apart from Chibnall's insulin project, which was funded by the pharmaceutical firms ICI and Eli Lilly (de Chadarevian, 1996; Gay, 2007; Weatherall and Kamminga, 1992).
[42] Martin and Synge, 1941; Consden, Gordon and Martin, 1944; Sanger, 1945, 1949a, 1988; Gordon, 1977; Smith, 1977; Pedersen, 1983. Research stays in Tiselius's laboratory learning separation methods were common initiatives in biochemistry from the 1930s onwards and resulted in multiple transfers of techniques from Uppsala to Cambridge and other European and North-American institutions (e.g. Kay, 1988). During his stay and encounter with Synge, Sanger became familiar with ionophoresis, another biochemical separation technique based on the electric charge of amino acids. Sanger applied this method to his late insulin experiments and determination of RNA sequences during the 1960s. His first experiments with ionophoresis in his notebooks are labeled as having been conducted "in Uppsala" and bear the initials "RLMS," referring to Richard [Laurence Millington] Synge (Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Book 3, p. 494 and Uppsala Book, reference number SA/BIO/P/1/6).
[43] Quotes from Gordon, Martin and Synge, 1943; Consden, Gordon and Martin, 1944, p. 224; Consden, Gordon, Martin and Synge, 1947, p. 1947.
[44] On gramicidin: Sanger, 1946. See also Sanger, 1951a, p. 463; 1953a, p. 353 and Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Book 4, quote from p. 681. In his study of Sanger's early methods, Ross has claimed that he provided Martin and Synge with DNFB-labeled amino acids as controls for their gramicidin project (Ross, 2008).
[45] In his investigation of ways of knowing and working – or working knowledges – Pickstone shows how modern biomedicine after World War II and especially after the early 1970s reflected a renewed configuration of experimental, analytical and natural history traditions present in older disciplines and periods, such as 19th century synthetic chemistry (Pickstone, 2007, pp. 513-14). Genomics and the Human Genome Project may, in this regard, be seen nowadays as framed in natural history and analysis – i.e. classificatory and routine endeavors – despite their necessary techniques having involved "much experimentation" (ibid., 2001, pp. 2-3 and 9). Sanger's early career represented such an initially experimental nature of sequence determination. With its spread and particularly its application, the other descriptive working knowledges of this practice rose, becoming an increasingly repetitive and routine activity. Sequence determination, in this regard, represents a particularly hybrid practice involving, on

the one hand, the performance of experiments and the generation of new knowledge, but, on the other, being a repetitive activity yielding always the same results through the routine application of methods (García-Sancho, 2008, pp. 37 and ff). Sequence determination, thus, shows how inappropriate categories such as science, technology and interdisciplinarity are for historically analyzing the 20th century, as some historians have argued (Edgerton, 1999, 2006; see note 70).

[46] Nevertheless, according to Fischer, chemistry and biology had increasingly followed separate paths during the second half of the 19th century and at that time (1907) were beginning to converge again (Fischer, 1907). In his classical history of biochemistry, Marcel Florkin has portrayed a similar divorce and reunification of both disciplines (Florkin, 1972, vol. 30, ch. 12 and ff.).

[47] Quotes from Fischer, 1907, p. 1765. The development of analytical and synthetic chemistry during the 19th century is the object of an ongoing PhD dissertation by Catherine Jackson at the Department of Science and Technology Studies of University College London (Jackson, 2007).

[48] Quotes from Fischer, 1907, pp. 1761 and 1753; Fruton, 1985, pp. 326-30.

[49] Fischer, 1907, p. 1761.

[50] Chibnall, 1942, p. 137. Bergmann edited the last volumes of the collected works of Fischer, published in a series of eight issues, four of which appeared after his death (Fruton, 1985, note 30). One of them, offered by Fischer's son, accompanies Sanger's laboratory notebooks (see note 14).

[51] Kohler, 1982, p. 7.

[52] A researcher happy with this change of orientation was Pirie, Sanger's first PhD Supervisor, who had complained during Hopkins's Chair for the lack of chemical contents in the department's teaching program (Kohler, 1982, p. 84).

[53] Kohler, 1982, p. 89; Synge and Williams, 1992, p. 81; Weatherall and Kamminga, 1992, pp. 83-84.

[54] Sanger's framing in the parameters of Chibnall's group is also reflected in his little interactive attitude towards crystallography. His contacts with crystallographers did not achieve "the experimental level," as Sanger has acknowledged in interviews (Sanger, 1988, 1992). Particularly remarkable in this regard was the connection between Sanger and Dorothy Hodgkin, an Oxford-based crystallographer working on insulin at the same time he was developing his sequence determination methods. Hodgkin has complained of occasional indifference of Sanger (Ferry, 1998, pp. 327-28) and the only evidence of contacts in Sanger's laboratory notebooks is a letter dated August 1945 attached to the experiments. In it, Hodgkin and members of her group send the results of an x-ray analysis of a sample of the protein gramicidin S previously delivered by Sanger (Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Book 2, p. 232, attached letter; see notes 15 and 22). The relationship between Sanger and Hodgkin, and more generally between crystallography and protein chemistry in insulin analysis, is an avenue for further research.

[55] Edman, 1950, pp. 284 and 292.

[56] Quotes from Hirs, Moore and Stein, 1960, p. 633; Stein and Moore, 1972, p. 85; see also Spackman, Stein and Moore, 1958. The preference of Edman's technique over Sanger's was partially due to the visit of the Swedish researcher to the Rockefeller Institute in the late 1940s. There, he interacted with the members of Bergmann's school, including Stein and Moore. Edman himself was involved in the design of an automatic apparatus performing all the steps of sequence determination and not only amino acid analysis during the late 1960s. In the following decade, there were a number of commercial machines of this sort available, namely one marketed by the firm Beckman Instruments (García-Sancho, 2008, ch. 3; Hartley, 1970, pp. 31 and further).

[57] Sanger, 1988, p. 12. De Chadarevian has suggested certain rivalry between Sanger, and Stein and Moore (1996, p. 371, note 36). This may be true, especially given the linkage of the American researchers with Bergmann and the disagreements between the latter and Chibnall's group over protein structure (see Section 1.2). Sanger has denied this, claiming that Stein and Moore were good friends of his, despite having different approaches to sequence determination (Sanger, personal communication, 2005).

[58] Sanger, Hartley et al, 1959; Sanger and Milstein, 1961; Sanger, 1963.

[59] Sanger, 1988, p. 11. Sanger had originally shown resistance towards this technique, according to him because of suspicion that the sequence of the fragments could be altered after the labeling (Sanger, 1988). His reluctance also relates to Sanger's dislike of physics, the discipline from which radioactive labeling came (Creager and Santesmases, eds., 2006). Since his undergraduate years, Sanger had avoided this discipline, in which he always obtained poor results (Sanger, 1992). This dislike also explains, partially, his alleged indifference towards crystallography, a technique with which he seldom interacted, despite also seeking the structure of proteins (see note 29).

[60] Creager and Santesmases, eds., 2006; Holmes, 2001, pp. 57-60; id. 2007, pp. 154-59.

[61] Sanger, 1988, pp. 11-14; 1955; Sanger, Hartley et al, 1959; Sanger and Milstein, 1961; Sanger, 1963.

[62] Ross, 2008; Strasser, 2008. Another influence in Sanger's espousal of comparative methods was his father, Frederick Sanger, a physician who cooperated with immunologist George Nuttall in the comparison of blood sera in Cambridge. He died fairly young, before the start of his son's PhD studies (Strasser, 2008; Sanger, 1992).
[63] Sarkar, 1992; 1998, ch. 6. Sarkar distinguishes various types of reductionism shaping biological explanation since classical genetics in the early 20th century. The one characteristic of molecular biology is physical reductionism, in which the connection between genotype and phenotype is established in terms of the chemical and physical interactions between the components of DNA, understood as a macromolecule (Sarkar, 1998, pp. 136-37).
[64] Sanger, 1988, p.11; 1987a; 1992. This situation contrasted with the results produced by Tuppy – Sanger's former assistant – and Anfinsen in protein sequence comparison (Strasser, 2008). The latter introduced these comparative practices into evolutionary research through *The Molecular Basis of Evolution*, a book published in 1959 (Ross, 2008). One of the reasons for Sanger's stagnation was the specific structure and composition of insulin, which made it unsuitable for comparisons (Strasser, personal communication).
[65] De Chadarevian, 2002, pp. 61 and ff.
[66] Sanger, 1987a; 2005. Other reasons in which Sanger has justified his move to the LMB are administrative in nature, namely the Medical Research Council being his funding body – as in the new center – and his not wanting to teach – the LMB researchers were exempted from such duty (Sanger, 2005, 1992). The relative importance of all these factors in Sanger's move will not be directly addressed in this paper which, in line with de Chadarevian (1996), will frame this move in a more general shift from biochemistry to molecular biology. I will, nevertheless, investigate in more detail the impact of sequence determination on molecular biology and of molecular biology on sequence determination as a practice (see below).
[67] For autobiographical accounts on the emergence of molecular biology see Crick, 1988; Brenner 2001. My proposal of a shift in Sanger's sequence determination from chemistry to biology after his move to the LMB links this paper to the project *The Nature of Evidence: How Well Do Facts Travel?*, recently concluded at the London School of Economics. This project included a line specifically dealing with biological facts conducted by Sabina Leonelli, currently based in the ESRC Centre for Genomics in Society, University of Exeter (Leonelli, 2008). See http://www.lse.ac.uk/collections/economicHistory/Research/facts/ (last accessed January 2009).
[68] De Chadarevian, 1996, 1999, 2002, ch. 7; Crick, 1988, Sanger, 1988.
[69] Olby, 1992 [1974]; Stent, 1968; Morange, 2000; de Chadarevian, 2002; Sarkar, 1998; Abir-Am, 1982, 1992.
[70] De Chadarevian, 2002, 1996. This transformation is exemplified by Marshall Nirenberg and Heinrich Matthaei, biochemists at the National Institutes of Health and discoverers of the mechanism of the genetic code in 1961. Lily Kay's investigation into this achievement also portrays the departure of molecular biology from physics and its incorporation of biochemical practices during the mid-late 50s (Kay, 2000, especially ch. 6).
[71] Olby, 1992 [1974]; Watson and Crick, 1953ab)Crick, 1958; de Chadarevian, 1996, pp, 375-376.
[72] Crick, 1958, pp. 138-63; Kay, 2000, ch. 4.
[73] De Chadarevian, 1996, pp. 379-82; 1999, p. 204; Ingram, 1956. Another important cooperation for this achievement was that between Ingram and Hermann Lehmann, a physician and medical researcher who collected hemoglobin variants all around the world and distributed them upon request among biologists and doctors. X-ray crystallographers Perutz and Kendrew were also studying the three-dimensional structure of hemoglobin at the Cavendish (de Chadarevian, 1998, pp. 178 and ff.).
[74] De Chadarevian, 1996, 2002.
[75] Sanger, 2005; 1988, p. 14; Crick, 1988, pp. 102-107. The lectures were given by Crick, Brenner and Seymour Benzer, another self-declared molecular biologist visiting Cambridge from the United States at that time. Benzer and Ingram simultaneously taught Lehmann the protein fingerprinting techniques and cooperated with this scientist in the characterization of hemoglobin variant D (de Chadatevian, 1998. p. 186; see previous note).
[76] Sanger, 2005, 1992, 2005, see note 36. Sanger's attitude towards professional identity contrasts with other contemporary scientists and is a key factor in the redefinition of his sequencing techniques. On the one hand, he was not hostile to molecular biology as some biochemists between the 1950s and 70s were (e.g. Chargaff, 1978). On the other, he remained indifferent towards the expansion of the new discipline, unlike more enthusiastic colleagues such as Severo Ochoa (Santesmases, 2002, see note 66). On molecular biology, biochemistry and professional identities see Abir-Am, 1992 and de Chadarevian and Gaudilliere, eds., 1996.

[77] De Chadarevian, 1996, p. 385.
[78] Quote from de Chadarevian, 1996, p. 380; Crick, 1988, pp. 102-107; Sanger, 1988, 2005.
[79] Crick and Brenner's personalities contrast with that of Smith, the researcher Sanger cites as having introduced him into the problems of DNA, RNA and protein synthesis (see above). Smith, a method-oriented scientist reluctant to visibility, was more similar to Sanger, both of them being pioneers in the application of paper chromatography to nucleic acids (Bretscher, 2003).
[80] Autobiographical accounts: Crick, 1988, Brenner, 2001; Watson, 1969, 2003. Popular and contemporary STS accounts: Judson, 1977, 1992; Wills, 1991; Cook-Deegan, 1994; Rose, 2006; Atkinson, Greenslade and Glasner, eds. 2007. The problem of the pervasiveness of retrospective accounts and of the view of a revolution is not exclusive of the historiography of molecular biology. Thomas Lean has shown that autobiographies of computer scientists and the application to history of Moore's Law have led to the idea of inevitability when talking about the progress of computing (Lean, 2008). Equally, Pnina Abir-Am and Clark Elliot have raised similar problems in anniversaries or commemorations of different scientific disciplines, among them molecular biology (Abir-Am and Elliot, eds., 1999, Abir-Am 1999).
[81] Quote from de Chadarevian and Kamminga eds., 1998, p. 1. See also Abir-Am, 1982; Kay, 1992; Olby, 1990; Fox Keller, 2000.
[82] Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin Book 12, pp. 2377 and ff, and Insulin Book 13; RNA and DNA books, files number SA/BIO/P/1/20 and ff. See also Sanger, 1988, pp. 102 and ff.
[83] Florkin, 1977, vol. 32; Gaudillière, 1992; Strasser, 2006b; Rheinberger, 1993; 1997, esp. ch. 10.
[84] Crick, 1958, p. 153.
[85] De Chadarevian and Kamminga, eds., 1998, ch. 1.
[86] Sanger, 1988, pp. 14-20.
[87] Sanger, Brownlee and Barrell, 1965; Brownlee, Sanger and Barrell, 1968. The attempts outside the LMB, as well as Sanger's RNA methods, are being investigated by Jerôme Pierrel in an ongoing PhD dissertation at the University Louis Pasteur in Strasbourg. Apart from Sanger's work, Pierrel has analyzed Walter Fiers's determination of the RNA virus MS2, first organism with a completed sequence in 1976 at the University of Ghent (Pierrel, 2008; Fiers et al, 1976). Due to the completeness and quality of Pierrel's work, this paper will not explore in more detail Sanger's RNA techniques.
[88] Sanger, 1988, pp. 20-21; 1980, p. 431; Sanger and Dowding, 1996, pp. 339-344; Fruton, 1999, pp. 412-13.
[89] Sanger, 1980, pp. 432-437; Wills, 1991, pp. 40-45. In the plus and minus method, Sanger achieved the selective stop of polymerase by including as loose nucleotides to be incorporated to the DNA either the same type of base (adenine, cytosine, guanine or thymine) or three of them, removing the other. This way, the enzyme would always stop at the only nucleotide included (plus method) or before the removed one (minus method) (Sanger and Coulson, 1975a, pp. 441-448; Sanger, 1975b, pp. 324-28). In the second technique – dideoxy method – normal nucleotides were combined with chemically modified ones (dideoxy) able to stop the polymerase reaction at each of the four bases. This permitted the achievement of DNA fragments which finished alternatively at dideoxiadenine, dideoxicytosine, dideoxiguanine and dideoxithymine (Sanger, Nicklen and Coulson, 1977, pp. 5463-5467; Garesse, 1987, pp. 72-74; id., 1994, pp. 77-81).
[90] Sanger, 1988, pp. 22-24; Wills, 1991, pp. 40-45. Gel electrophoresis could be two-dimensional as much as paper chromatography could be one-dimensional. It all depended on whether the performed separation was double – first vertical and then horizontal – or single. The use of electrophoresis in DNA sequence determination at Sanger's group gradually shifted from two-dimensional to one-dimensional (see below).
[91] Quote from Sanger, 2005; see also id., 1988; 1987a; 1992.
[92] Sanger, 2005.
[93] Sanger, 1988, p. 22.
[94] Sanger's assistants have also stressed the importance of these failed attempts and the "determination" of their boss in developing suitable techniques (Coulson, 2005).
[95] Sanger's Laboratory Notebooks, Wellcome Trust, London, Insulin, RNA and DNA Books (e.g. Insulin Book 10, p. 2140; RNA Books, folder number SA/BIO/P/1/21, experiment number R45; DNA Books, file number SA/BIO/P/1/42, experiment number D80).
[96] Id., DNA Books, files number SA/BIO/P/1/40 to SA/BIO/P/1/41. See also Murray, 1970; Robertson, Barrell, Weith and Donelson, 1973; Sanger, Donelson, Coulson, et al, 1973.
[97] Ibid., file number SA/BIO/P/1/43, experiments number D99(7) and ff., also files number SA/BIO/P/1/31 to SA/BIO/P/1/34.

[98] Holmes has applied this concept to the careers of Meselson, Stahl and Benzer, other key figures in the early development of molecular biology. His methodology also involved interviews and detailed analyses of notebooks and published papers, in order to explore how scientific trajectories evolve in a day-to-day basis. Holmes shares the same aim as this paper: to determine the motivations leading to scientific choices and to challenge retrospective accounts which reduce them to Eureka moments (Holmes, 2001, esp. pp. 3 and ff., 2006, esp. pp. VIII-XIX and ch. 6, pp. 220-21). On history and collective memory of scientists more generally see Abir-Am and Elliot, 1999.

[99] The necessity of this collective historiography was first raised by Adam Bostanci, researcher working on more recent developments in DNA sequence determination (Bostanci, 2004, 2005). There is, certainly, not a collective study, as the only available investigations focus either on Sanger (de Chadarevian, 1996, 1999; this paper) or on other US researchers such as Ray Wu (Onaga, 2005; see below).

[100] The polymerase used by Sanger's group was, nevertheless, chemically modified, in order to make it suitable for the copying procedure (Coulson, personal communication, 2005).

[101] Wu and Kaiser, 1968; Wu and Taylor, 1971)Billeter, Weissmann et al, 1969.

[102] Wu and Kaiser, 1968, pp. 528-34; Wu and Taylor, 1971, pp. 496-509; Billeter and Weissmann, 1969, pp. 1085-86. In a retrospective account, Wu has stressed his group's role in developing the copying approach to DNA sequence determination, usually attributed solely to Sanger (Wu, 1994).

[103] Thurtle, 1998; Kay, 1988; Pederson, 1983; Gordon, 1977.

[104] E.g. Dingman and Peacock, 1968a; 1968b. Historian Howard Chiang has shown the different identities that electrophoresis was given from its invention in the 1930s up to the 60s, when it began being widely used by molecular biologists. The technique was originally invented as an instrument creating a "moving boundary" in which the separation of the molecules was not complete. During the 1940s and 50s, researchers in medical biochemistry and protein chemistry tested different separation media – filter paper, starch grain and agar gel – and created the concept of "zone electrophoresis," in which the separation of the molecules was differentiated in independent areas on the surface. Electrophoresis, Chiang concludes, acquired its current identity of "molecular-sieving" technique in the early 1960s, when polyacrylamide gel emerged as the preferred separation medium (Chiang, 2007).

[105] Sanger's different use of gel electrophoresis meant that its application to sequence determination and the interpretation of the results required certain competencies. Rafael Garesse, a Spanish molecular biologist and postdoctoral fellow at the LMB during the early 1980s, has noted that both the preparation of the gels and the visual analysis of the bands to determine the sequence necessitated "much training" and were correctly achieved only "after a few attempts" (Garesse, 2005). Sequence determination, consequently, was becoming a practice requiring specific skills only available at certain laboratories. During the 1950s and 60s, the use of paper chromatography and other components of Sanger's protein fingerprinting technique had also proved difficult for evolutionary biologists incorporating sequence determination (Ross, 2008, pp. 12 and ff.).

[106] Sanger and Coulson, 1975; Sanger, Nicklen and Coulson, 1977. *Sequencing* was also used by Wu in one of his sticky ends papers in 1971 (Wu and Taylor, 1971, p. 491), but not systematically incorporated to published literature until Sanger's 1975 and 77 articles.

[107] De Chadarevian, 1996, 2002, esp. ch. 8.

[108] The will of molecular biologists to move towards new research horizons is reflected in a 1963 letter of Brenner to Perutz, then Director of the LMB. In it, Brenner stated that "all the *classical* problems of molecular biology" had "either been solved" or would be solved "in the next decade," and it was, therefore, necessary to extend investigations "to other fields of biology." Brenner would base on that letter his further 1970s research on the worm *C. elegans* (Brenner, 1963, p. X; García-Sancho, 2008, ch. 2).

[109] Brenner had already cooperated with Jacob in the discovery of messenger RNA, an achievement related to the investigations on the genetic code during the late 1950s and early 60s (Brenner, Jacob and Meselson, 1961). The decipherment of the code's mechanism by Nirenberg and Matthaei was based on the synthesis of messenger RNA in the test tube and did not incorporate sequencing techniques (Kay, 2000, chs. 5 and 6; Matthaei and Nirenberg, 1961; Nirenberg and Matthaei, 1961).

[110] The differences in approach between biology and chemistry have also been noted by biochemist and Nobel Prize winner Arthur Kornberg, who defined them as "two cultures" evoking the gulf that C.P. Snow postulated between the humanities and natural sciences (Kornberg, 1987, p. 6888; García-Sancho, 2008, pp. 198 and ff.).

[111] Brenner, 2001, pp. 108-115.

[112] Sanger's Laboratory Notebooks, Wellcome Trust, London, DNA Books, file number SA/BIO/P/1/43, quote from experiment D93. Coulson, 2005.

[113] Sanger introduced the term *information* to describe the potentialities of the DNA sequence. This concept, central in the development of molecular biology (Kay, 2000; Sarkar, 1996; Fox Keller, 1995, ch. 3; Brandt, 2005), was widely used by Brenner, Crick and other LMB researchers. Whereas Sanger's protein techniques were compared with a jigsaw-puzzle game – i.e. assembling overlapping fragments – in his DNA papers Sanger began talking about sequences that could be "read off" (Sanger, 1975a p. 443; 1977, p. 5463; García-Sancho, 2007a, pp. 21-24; 2007b). The role of DNA in the development of an organism from embryo to adult – also referred by Sanger – was a crucial concern in Brenner's investigations on the worm *C. elegans*, conducted at the same time of the invention of the plus and minus, and dideoxy methods (id., 2008, ch. 2).
[114] Sanger, 1975b, p. 317.
[115] Sanger, 2005.
[116] Sanger and Coulson, 1975a, p. 443; id., 1977, p. 5463.
[117] In a recent biography of Elizabeth Blackburn – PhD student at Sanger's laboratory – Catherine Brady has claimed that ØX-174 was suggested to the group by postdoctoral fellow John Sedat (Brady, 2007, p. 27).
[118] Sanger's immersion in molecular biology was common among biochemists between the 1960s and 70s. María Jesús Santesmases has shown the incorporation of "genetic thinking" into Ochoa's work when he was shifting his research agenda from enzymology to the genetic code in the early 60s (Santesmases, 2002, pp. 193-94). Other researchers, on the contrary, preferred to maintain their biochemical identity (Chargaff, 1978; Mullis, 1998; García-Sancho, 2008, pp. 193-205; see note 41).
[119] Gilbert, 1980, p. 409; Cook-Deegan, 1994, pp. 61-64; Sutcliffe, 1995.
[120] Maxam and Gilbert, 1977; Gilbert, 1980; Soutcliffe, 1995; Sanger and Dowding, 1996, pp. 343-44; Wills, 1991, pp. 45-47.
[121] Cook-Deegan, 1994, pp. 64-77; García-Sancho, 2008, ch. 3. Sanger's technique limitation was partially solved in the late 1970s, through the cloning of the DNA to be sequenced (normally double-stranded) in a single-stranded bacteriophage. This way, it was possible to obtain single-stranded DNA from every template (Sanger, 1988, pp. 23-24; id., 1980, pp. 437-39). The dideoxy method could not be directly applied to double-stranded DNA until the mid 1980s, with the advent of denaturation techniques for separating the DNA strands (Garesse, personal communication, 2005).
[122] Sanger, 2005; Coulson, 2005.
[123] E.g. Garesse, 1987, 1994. This situation points to another extra-technological factor in the development of sequencing: the specifics of DNA. The capacity of this molecule for expressing and duplicating itself made that first molecular biologists and then Sanger incorporated those mechanisms, considering them the natural and elegant way of sequencing and researching on DNA. Proteins and RNA lacked this capacity of self-replication and needed to be investigated through other procedures, such as chemical degradation.
[124] García-Sancho, 2008, chs. 2-3. The field of protein sequence determination witnessed the opposite process: Edman's method was preferred over Sanger's since the late 1950s. This may, equally, be due to Edman's technique being better adapted to the working procedures of biochemistry. Edman had previously developed links with the Rockefeller Institute – a leading center in this field – and, crucially, with Stein and Moore, the initiators of the automation of protein sequence determination (see note 30).
[125] The demarcation between practices and disciplines has also consequences for the general historiography of science. This distinction is the basis for John Pickstone's "new history of science, technology and medicine," based on interacting "ways of knowing" and "working" which after the 19th century do not correspond with disciplinary boundaries (Pickstone, 2001, 2007, see note 5). David Edgerton has equally noted that the separation between science and technology, as well as categories such as interdisciplinarity, are no longer tenable for analyzing historically the 20th century (Edgerton, 1999, 2006).
[126] The impact of revolutionary rhetoric on current biomedical research has been analyzed by Jane Calvert and Joan Fujimura when investigating systems biology: researchers have constructed the identity of this new field by defining it as a "revolution" with regard to genomics. Systems biology as a field arose barely a decade after genomics (Calvert and Fujimura, 2007).

REFERENCES

-Abir-Am P. (1982) “The discourse of physical power and biological knowledge in the 1930s: a reappraisal of the Rockefeller Foundation ‘policy’ in molecular biology” in *Social Studies of Science*, 12.

-Abir-Am P. (1992) “The politics of macromolecules: molecular biologists, biochemists and rhetoric” in *Osiris*, Second Series, 7.

-Abir-Am P. (1999) “The first American and French commemorations in molecular biology: from collective memory to comparative history” in *Osiris, Second Series*, 14.

-Abir-Am P. and Elliot C.A. (eds.) (1999) “Commemorative practices in science: historical perspectives on the politics of collective memory,” special issue of *Osiris, Second Series*, 14.

-Armon R. (2006) “Joseph Needham, the mechanist-vitalist debate, and the origins of biochemistry, 1922-42,” *Scientists and Social Commitment* meeting, Science Museum, London.

-Astbury W. (1934) “X-ray studies of protein structure” in *Cold Spring Harbor Symposia on Quantitative Biology*, 2, pp. 15-27.

-Atkinson P., Greenslade H. and Glasner P. (eds.) (2007) *New Genetics, New Identities* (Routledge).

-Billeter M.A. and Weissmann C. (1969) “Sequence of the first 175 nucleotides from the 5’ terminus of Qβ RNA synthesised in vitro” in *Nature*, 224, pp. 1083-86.

-Bonneuil C. and Gaudillière J.P. (2007) "Navigating the post-Fordist DNA: network regulations and variability in genomics and society" International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB) Biennial Meeting, University of Exeter.

-Bostanci A. (2004) “Sequencing human genomes” in J.P. Gaudillière and H.J. Rheinberger (eds.) *From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth Century Biology* (London: Routledge).

-Bostanci, A. (2005) “Response to García-Sancho,” seminar at the ESRC Centre for Genomics in Society (Egenis), University of Exeter (UK).

-Brady C. (2007) *Elizabeth Blackburn and the Story of Telomers* (MIT).

-Brandt C. (2005) “Genetic code, text, and scripture: Metaphors and narration in German molecular biology” in *Science in Context*, 18(4).

-Brenner S., Jacob F. and Meselson M. (1961) “Unstable intermediate carrying information from genes to ribosomes for protein synthesis” in *Nature*, 190, pp. 576-581.

-Brenner S. (1963) “Letter to Max Perutz” in W. Wood (ed.) (1988) *The nematode C. Elegans* (New York: Cold Spring Harbor).

-Brenner S. (2001) *My Life in Science* (London: BioMed), especially chs. 7-8.

-Bretscher M.S. (2003) “Obituary: John Smith” in *The Independent*, November 27th.

-Calvert J. and Fujimura J. (2007) “Systems biology: the revolution after the revolution?,” International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB) biennial meeting.

-Chargaff E. (1978) *Heraclitean Fire: Sketches from a Life before Nature* (Rockefeller University Press).

-Chiang H. (2007) “Separating molecules, building biology: the evolution of electrophoretic instrumentation and the material epistemology of molecular biology, 1945-1965,” Biennial Meeting of the International Society for the History, Philosophy and Social Studies of Biology, University of Exeter.

-Chibnall C. (1942) “Amino-acid analysis and the structure of proteins” in *Proceedings of the Royal Society of London*, series B, 131, pp. 136-160.

-Clarke H. (1944) “Obituary: Max Bergmann, 1886-1944)” in *Science*, 102, pp. 168-70.

-Consden R., Gordon A., Martin A. (1944) “Qualitative analysis of proteins: a partition chromatographic method using paper” in *Biochemical Journal*, 38, pp. 224-32.

-Consden R., Gordon A., Martin A., Synge R. (1947) “Gramicidin S: the sequence of the amino-acid residues” in *Biochemical Journal*, 41, pp. 596-602.

-Cook-Deegan R. (1994) *The Gene Wars: Science, Politics and the Human Genome* (London: W.W. Norton and Company).

-Coulson A. (2005) Interview with Miguel García-Sancho, Laboratory of Molecular Biology, Cambridge.

-Creager A. (1998) “Producing molecular therapeutics from human blood: Edwin Cohn’s wartime enterprise” in de Chadarevian S. and Kamminga H. (eds.) (1998) *Molecularizing Biology and Medicine: New Practices and Alliances, 1910s-1970s* (Harwood).

-Creager A. and Santesmases M.J. (eds.) (2006) *Radiobiology in the Atomic Age: changing research practices and policies in comparative perspective*, special issue of the *Journal of the History of Biology*, 39.

-Crick F. (1958) “On protein synthesis” in *Symposium of the Society for Experimental Biology*, 12, pp. 138-163.

-Crick F. (1988) *What Mad Pursuit: A Personal View of Scientific Discovery* (New York: Basic Books), pp. 102-107.

-de Chadarevian S. (1996) "Sequences, conformation, information: biochemists and molecular biologists in the 1950s" in *Journal of the History of Biology*, 29.

-de Chadarevian S. (1998) "Following molecules: hemoglobin between the clinic and the laboratory" in id. and Kamminga H. (eds.) (1998) *Molecularizing Biology and Medicine: New Practices and Alliances, 1910s-1970s* (Harwood).

-de Chadarevian S. (1999) "Protein sequencing and the making of molecular genetics" in *Trends in Biochemical Sciences,* 24.

-de Chadarevian S. (2002) *Designs for Life: Molecular Biology after World War II* (Cambridge).

-de Chadarevian S. and Gaudillière J.P. (1996) "The tools of the discipline: biochemists and molecular biologists" in *Journal of the History of Biology*, 29.

-de Chadarevian S. and Kamminga H. (eds.) (1998) *Molecularizing Biology and Medicine: New Practices and Alliances, 1910s-1970s* (Harwood).

-Dingman C.W. and Peacock A.C. (1968) "Analytical studies on nuclear ribonucleic acid using polyacrylamide gel electrophoresis" in *Biochemistry,* 7(2), pp. 659-68.

-Donelson J. and Wu R. (1972) "Nucleotide sequence analysis of deoxyribonucleic acid" in *Journal of Biological Chemistry*, 247(14), pp. 4654-4668.

-Edgerton D. (1999) "From innovation to use: ten eclectic theses on the historiography of technology" in *History and Technology*, 16. French version available at *Annales HSS*, vols. 4-5, 1998.

-Edgerton D. (2006) *The Shock of the Old. Technology and Global History since 1900* (Oxford).

-Edman P. (1950) "Method for determination of the amino acid sequence in peptides" in *Acta Chemica Scandinavica*, 4, pp. 283-93.

-Edsall J.T. (1979) "The development of the physical chemistry of proteins, 1898-1940" in *Annals of the New York Academy of Sciences,* 325, special issue on *The Origins of Modern Biochemistry: A Retrospect on Proteins.*

-Ferry G. (1998) *Dorothy Hodgkin: A Life* (Granta), especially chs. 7-8.

-Fiers W. et al (1976) "Complete nucleotide-sequence of bacteriophage MS2-RNA: primary and secondary structure of replicase gene" in *Nature*, 260, pp. 500-507, 1976.

-Fischer E. (1902) "Über die Hydrolyse der Proteïnstoffe" in *Chemiker Zeitung*, 26.

-Fischer E. (1907) “Synthetical chemistry in its relation to biology” in *Transactions of the Chemical Society*, 91.

-Florkin M. (1972-1977) *A History of Biochemistry* in id. And Stotz E. (eds.) *Comprehensive Biochemistry* (Elsevier), especially vol. 30 and 32.

-Fox Keller E. (1995) *Refiguring life: changing metaphors in 20th century biology* (New York).

-Fox Keller E. (2000) *The Century of the Gene* (Harvard).

-Fruton J. (1972) *Molecules and Life: Historical Essays on the Interplay of Chemistry and Biology* (Wiley-Interscience).

-Fruton J. (1979) “Early theories of protein structure” in *Annals of the New York Academy of Sciences*, 325, special issue on *The Origins of Modern Biochemistry: A Retrospect on Proteins.*

-Fruton J. (1985) “Contrasts in scientific style. Emil Fischer and Franz Hofmeister: their research groups and their theory of protein structure” in *Proceedings of the American Philosophical Society*, 129(4).

-Fruton J. (1992) *A Skeptical Biochemist* (Harvard).

-Fruton J. (1999) *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology* (Yale).

-García-Sancho M. (2007a) “The rise and fall of the idea of genetic information (1948-2006)” in *Genomics, Society and Policy*, 2(3). Available free access at http://www.hss.ed.ac.uk/genomics/vol2no3/Garcia-sanchoabstract.htm

-García-Sancho M. (2007b) “Mapping and sequencing information: the social context for the genomics revolution.” *Endeavour*, 31(1), pp. 18-23. Available at http://dx.doi.org/10.1016/j.endeavour.2007.01.006

-García-Sancho M. (2008) *Sequencing as a Way of Work: A History of its Emergence and Mechanisation – From Proteins to DNA, 1945-2000*, PhD thesis, Centre for the History of Science, Imperial College, London.

-Gaudillière J.P. (1992) “J. Monod, S. Spiegelman et l’adaptation enzymatique: programmes de recherche, cultures locales et traditions disciplinaires” in *History and Philosophy of the Life Sciences*, 14.

-Gaudillière J.P. (2002) *Inventer la biomédicine: la France, l’Amérique et la production des savoirs du vivant (1945-1965)* (La Découverte).

-Garesse R. (1987) “Secuenciación de DNA y síntesis de oligonucleótidos” [“DNA sequencing and oligonucleotide synthesis”] in M. Vicente and J. Renart (eds.)

Ingeniería Genética [*Genetic Engineering*] (Madrid: Consejo Superior de Investigaciones Científicas).

-Garesse R. (1994) "Presente y futuro de las técnicas de secuenciación de DNA" ["Present and future of the DNA sequencing techniques"] in M. Vicente (ed.) *Avances en Ingeniería Genética* [*Advances in Genetic Engineering*] (Madrid: Consejo Superior de Investigaciones Científicas).

-Garesse, R. (2005) Interview with Miguel García-Sancho. "Alberto Sols" Institute for Biomedical Investigations, Madrid (in Spanish).

-Gay H. (2007) *The history of Imperial College London 1907-2007 : Higher Education and Research in Science, Technology and Medicine* (Imperial College Press).

-Gilbert W. (1980) "DNA sequencing and gene structure," Nobel Lecture, available at www.nobel.se/chemistry/laureates/1980/gilbert-lecture.html

-Gordon A.H., Martin A. and Synge R. (1943) "The amino acid composition of gramicidin" in *Biochemical Journal*, 37(1).

-Gordon A.H. (1979) "Electrophoresis and chromatography of amino acids and proteins" in *Annals of the New York Academy of Sciences*, 325, special issue on *The Origins of Modern Biochemistry: A Retrospect on Proteins.*

-Goven J. (2008) "Shaping the ELSA agenda: new technologies and old politics," International Conference of the UK Genomics Network and the Dutch Centre for Society and Genomics, Amsterdam.

-Hammond A. (2007) "J.B.S. Haldane and speciation: not a beanbag, but a full bag," *Annual Meeting of the British Society for the History of Science*, University of Manchester.

-Hartley B. (1970) "The primary structure of proteins" in T.W. Goodwin (ed.) *British Biochemistry: Past and Present*, Biochemical Society Symposia, nº 30 (London: Academic Press).

-Hartley B. (2004) "The first floor, Department of Biochemistry, University of Cambridge" in *IUBMB Life*, 56(7).

-Hirs C., Moore S., Stein W. (1960) "The sequence of the amino acid residues in performic acid-oxidized ribonuclease" in *Journal of Biological Chemistry*, 235(3), pp. 633-47.

-Hofmeister F. (1902) "Über Bau und Gruppierung der Eiweisskörper" in *Ergebnisse der Physiologie*, 1.

-Holmes F.L. (2001) *Meselson, Stahl and the Replication of DNA: A History of 'The Most Beautiful Experiment in Biology'* (Yale).

-Holmes F.L. (2006) *Reconceiving the Gene: Seymour Benzer's Adventures in Phage Genetics* (Yale).

-Ingram V. (1956) "A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobins" in *Nature,* 178.

-Jackson C. (2007) "Making substances speak: the public and private voices of Justus Liebig's organic chemistry," Annual Meeting of the British Society for the History of Science, University of Manchester.

-Jeppesen P., Barrell B., Sanger F. and Coulson A. (1972) "Nucleotide sequence of two fragments from the coat-protein cistron of bacteriophage R17 ribonucleic acid" in *Biochemical Journal,* 128 (5).

-Judson H.F. (1979) *The Eight Day of Creation: Makers of the Revolution in Biology* (Cape).

-Judson H.F. (1992) "A History of the Science and Technology Behind Gene Mapping and Sequencing" in D.J. Kevles and L. Hood (eds.) *The Code of Codes: Scientific and Social Issues in the Human Genome Project* (Harvard).

-Kay L. (1988) "Laboratory technology and biological knowledge: the Tiselius electrophoresis apparatus, 1930-1945" in *History and Philosophy of the Life Sciences,* 10(1).

-Kay, L. (1993) *The Molecular Vision of Life: Caltech, the Rockefeller Foundation and the Rise of the New Biology* (Oxford).

-Kay L. (2000) *Who Wrote the Book of Life: A History of the Genetic Code* (Stanford).

-Kevles D.J. and Hood L. (eds.) 1992. *The Code of Codes: Scientific and Social Issues in the Human Genome Project* (Harvard University Press).

-Kohler R. (1982) *From Medical Chemistry to Biochemistry* (Cambridge).

-Kornberg A. (1987) "The two cultures: chemistry and biology" in *Biochemistry*, 26, pp. 6888-91.

-Lagnado J. (2005) "Questions and answers with Fred Sanger" in *The Biochemist.*

-Lean T. (2008) "In just a day you'll be talking to it like an old friend: the development of the 1980s home microcomputer in Britain," Lunchtime Seminar Series, Centre for the History of Science, University of Manchester.

-Leonelli, S. (2008) "Circulating evidence across research contexts: the locality of data and claims in model organism research," working paper, project *The Nature of Evidence: How Well do Facts Travel,* London School of Economics. Available at http://www.lse.ac.uk/collections/economicHistory/pdf/FACTSPDF/HowWellDoFactsTravelWP.htm

-Martin A. (1952) “The development of partition chromatography,” Nobel Lecture, available at www.nobel.se/chemistry/laureates/1952/martin-lecture.html

-Martin A. and Synge R. (1941) “A new form of chromatogram employing two liquid phases” in *Biochemical Journal*, 35, pp. 1358-68.

-Matthaei H. and Nirenberg M. (1961) “Characterisation and stabilisation of DNAase-sensitive protein synthesis in *E. Coli* extracts” in *Proceedings of the National Academy of Sciences*, 47, pp. 1580-88.

-Maxam A. and Gilbert W. (1977) “A new method for sequencing DNA” in *Proceedings of the National Academy of Sciences*, 74.

-Meselson M. and Stahl F. (1958) “The replication of DNA in *Escherichia Coli*” in *Proceedings of the National Academy of Sciences*, 44.

-Moore S. and Stein W. (1972) “The chemical structure of pancreatic ribonuclease and deoxiribonuclease,” Nobel Lecture, available at www.nobel.se/chemistry/laureates/1972/moore-lecture.html

-Morange M. (2000) *A History of Molecular Biology* (Harvard).

-Morris P. (1998) “Chromatograph” in R. Bud (ed.) *Instruments of Science: An Historical Encyclopaedia* (London: Science Museum).

-Mullis K. (1998) *Dancing Naked in the Mind Field* (Vintage).

-Murray K. (1970) “Nucleotide maps of digests of deoxyribonucleic acids” in *Biochemical Journal*, 118, pp. 831-41.

-Neuberger A. (1939) “Chemical criticism of the cyclol and frequency hypothesis of protein structure” in *Proceedings of the Royal Society of London*, B series, 127, pp. 25-26.

-Nirenberg M. and Matthaei H. (1961) “The dependence of cell-free protein synthesis in *E. Coli* upon naturally occurring or synthetic polyribonucleotides” in *Proceedings of the National Academy of Sciences*, 47, pp. 1588-1602.

-Olby R. (1992 [1974]) *The Path to the Double Helix: The Discovery of DNA* (Courier Dover Publications).

-Olby R. (1990) “The molecular revolution in biology” in id. G. Cantor, J. Christie and J. Hodge (eds.) *Companion to the History of Modern Science* (Routledge).

-Onaga L. (2005) “Ray Wu and DNA sequencing,” paper presented at the ISHPSSB 2005 Meeting, University of Guelph.

-Paul D. (1984) “Eugenics and the left” in *Journal of the History of Ideas*, 45 (4).

-Peacock A.C. and Dingman, C.W. (1968) “Molecular weight estimation and separation of ribonucleic acids by electrophoresis in agarose-acrylamide composite gels” in *Biochemistry*, 7(2), pp. 668-74.

-Pederson K. (1983) “The Svedberg and Arne Tiselius. The early development of protein chemistry at Uppsala” in Semenza G. (ed.) *Selected Topics in the History of Biochemistry: Personal Recollections (Comprehensive Biochemistry,* vol. 35) (Elsevier).

-Pickstone J. (2001) *Ways of Knowing: A New History of Science, Technology and Medicine* (Chicago).

-Pickstone J. (2007) “Working knowledges before and after circa 1800: practices and disciplines in he history of science, technology, and medicine” in *Isis,* 98, pp. 489-516.

-Pierrel J. (2008) “Des RNAses à l’RNA: Walter Fiers” in *Bench-side story. RNA sequencing as a laboratory practice in the 1960s and 1970s in Cambridge, Ghent and Strasburg,* ongoing PhD dissertation, Centre for Interdisciplinary Research on Science and Technology, University Louis Pasteur, Strasbourg.

-Pirie W. (1939) “Amino-acid analysis and protein structure” in *Annual Report of the Chemical Society*, 36, pp. 351-353.

-Ramillon V. (2007) "The material economy of genomic research: automation, work division and productivity,” International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB) Biennial Meeting, University of Exeter.

-Rheinberger H.J. (1993) “Experiment and orientation: early systems of in vitro protein synthesis” in *Journal of the History of Biology*, 26.

-Rheinberger H.J. (1997) *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford).

-Rheinberger H.J. and Gaudillière J.P (eds.) (2004a) *Classical Genetic Research and its Legacy: The Mapping Cultures of 20th Century Biology* (London: Routledge).

-Rheinberger H.J. and Gaudillière J.P (eds.) (2004b) *From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth Century Biology* (London: Routledge).

-Robertson H.D., Barrell B., Weith H., Donelson J. (1973) “Isolation and sequence analysis of a ribosome-protected fragment from bacteriophage ØX 174 DNA” in *Nature New Biology*, 241, pp. 38-40.

- Rose N. (2006) *The Politics of Life Itself: Biomedicine, Power and Subjectivity in the Twenty-First Century* (Princeton).

-Ross S. (2008) “How sequences became simple: proteins and evolution before molecular evolution,” workshop *Making Sequences Matter,* University of Yale.

-Sanger F. (1945) “The free amino groups of insulin” in *Biochemical Journal*, 39, pp. 507-515.

-Sanger F. (1946) “The free amino group of gramicidin S” in *Biochemical Journal*, 40 (2).

-Sanger F. (1949a) “The terminal peptides of insulin” in *Biochemical Journal*, 45.

-Sanger F. (1949b) “Some chemical investigations on the structure of insulin” in *Cold Spring Harbor Symposia on Quantitative Biology: Amino Acids and Proteins*, 14.

-Sanger F. (1959) “Chemistry of insulin” in *Science*, 129, pp. 1340-44.

-Sanger F. (1963) “Amino acid sequences in the active centres of certain enzymes” in *Proceedings of the Chemical Society (UK)*, March 1963.

-Sanger F. and Coulson A. (1975a) “A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase” in *Journal of Molecular Biology*, 94.

-Sanger F. (1975b) “The Croonian Lecture” in *Proceedings of the Royal Society of London*, 191, B series.

-Sanger F. (1980) “Determination of nucleotide sequences in DNA,” Nobel Lecture, available at www.nobel.se/chemistry/laureates/1980/sanger-lecture.html

-Sanger F. (1987a) Interview with Horace F. Judson, Biochemical Society, London. Video available through the Biochemical Society.
-Sanger F. (1987b) “Reading the messages in our genes” in *New Scientist*, 21st March.

-Sanger F. (1988) “Sequences, sequences and sequences” in *Annual Review of Biochemistry*, 57.

-Sanger F. (1992) Interview with G.G. Brownlee, Imperial College, London. Video available through the Biochemical Society.

-Sanger F. (2001) “The early days of DNA sequences” in *Nature Medicine*, 7(3).

-Sanger F. (2005) Interview with Miguel García-Sancho, Sanger Institute, Hinxton (Cambridge).

-Sanger F. and Dowding M. (eds.) (1996) *Selected Papers of Frederick Sanger (with Commentaries)* (London: World Scientific), pp. 7-8 and 339-344

-Sanger F. and Milstein C. (1961) “An amino acid sequence in the active centre of phosphoglucomutase” in *Biochemical Journal*, 79.

-Sanger F. and Thompson E.O. (1953a) “The amino-acid sequence in the glycyl chain of insulin: 2. The identification of lower peptides from partial hydrolysates” in *Biochemical Journal*, 53(3).

-Sanger F. and Thompson E.O. (1953b) “The amino-acid sequence in the glycyl chain of insulin: II. The investigation of peptides from enzymic hydrolysates” in *Biochemical Journal*, 53(3).

-Sanger F. and Tuppy H. (1951a) “The amino-acid sequence in the phenylalanyl chain of insulin: I. The identification of lower peptides from partial hydrolysates” in *Biochemical Journal*, 49(4).

-Sanger F. and Tuppy H. (1951b) “The amino-acid sequence in the phenylalanyl chain of insulin: II. The investigation of peptides from enzymic hydrolysates” in in *Biochemical Journal*, 49(4).

-Sanger F. et al (1955) “The disulphide bonds of insulin” in *Biochemical Journal,* 60(4).

-Sanger F. et al (1978) “The nucleotide sequence of bacteriophage ØX174” in *Journal of Molecular Biology*, 125.

-Sanger F. et al (1980) “Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing” in *Journal of Molecular Biology*, 143, pp. 161-78.

-Sanger F., Barrell B., Coulson A. (1977) “Nucleotide sequence of bacteriophage ØX-174” in *Nature*, 265.

-Sanger F., Brown H., Kitai R. (1955) “The structure of pig and sheep insulins” in *Biochemical Journal*, 60, pp. 556-564.

-Sanger F., Brownlee G.G., Barrell B. (1965) “A two-dimensional fractionation procedure for radioactive nucleotides” in *Journal of Molecular Biology*, 13(2), pp. 373-98.

-Sanger F., Brownlee G.G., Barrell B. (1968) “The sequence of 5s ribosomal ribonucleic acid” in *Journal of Molecular Biology*, 34(3), pp. 379-412.

-Sanger F., Donelson J.E., Coulson A. et al (1973) “Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA” in *Proceedings of the National Academy of Sciences*, 70(4), pp. 1209-13.

-Sanger F., Hartley B. and Naughton M. (1959) “The amino acid sequence around the reactive serine of elestase” in *Biochemical and Biophysical Acta*, 34.

-Sanger F., Nicklen S. and Coulson A. (1977) “DNA sequencing with chain terminating inhibitors” in *Proceedings of the National Academy of Sciences*, 74.

-Sanger, F., Coulson, A. et al (1982) “Nucleotide sequence of bacteriophage λ DNA” in *Journal of Molecular Biology*, 162 (4).

-Santesmases M.J. (2002) “Enzymology at the core: primers and templates in Severo Ochoa’s transition from biochemistry to molecular biology” in *History and Philosophy of the Life Sciences*, 24.

-Sarkar S. (1996) “Biological information: a sceptical look at some central dogmas in molecular biology” in id. (ed.) *The Philosophy and History of Molecular Biology: New Perspectives* (Dordrecht), pp. 187-233. Reprinted in id. (2005) *Molecular Models of Life: Philosophical Papers on Molecular Biology* (MIT).

-Sarkar S. (1998) *Genetics and Reductionism* (Cambridge).

-Sloan P.R. (ed.) *Controlling our Destinies: Historical, Philosophical, Ethical and Theological Perspectives on the Human Genome Project* (Indiana: University of Notre Dame).

-Smith E.L. (1979) “Amino acid sequences of proteins: the beginnings” in *Annals of the New York Academy of Sciences,* 325, special issue on *The Origins of Modern Biochemistry: A Retrospect on Proteins.*

-Spackman D., Stein W., Moore S. (1958) “Automatic recording apparatus for use in the chromatography of amino acids” in *Analytical Chemistry*, 30(7), pp. 1190-1207.

-Stent G.S. (1968) “That was the molecular biology that was” in *Science,* 160, pp. 390-95.

-Strasser, B. (2006a) “Collecting and experimenting: the moral economies of biological research, 1960s-1980s” in *History and Epistemology of Molecular Biology and Beyond: Problems and Perspectives* (Berlin: Max Planck Institute for the History of Science, preprint number 310).

-Strasser B. (2006b) “A world in one dimension: Linus Pauling, Francis Crick and the central dogma of molecular biology” in *History and Philosophy of the Life Sciences,* 28.

-Strasser (2008) “Collecting, comparing, computing: the natural history of molecular sequences, 1954-1966,” workshop *Writing Genomics: Historiographical Challenges for New Historical Developments,* Max Planck Institute for the History of Science, Berlin.

-Stretton A. (2002) “The first sequence: Fred Sanger and insulin” in *Genetics*, 162.

-Suárez E. (2007) "Evolutionary tools and comparative genomics: continuity in the shadow" in Biennial Meeting of the International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB), University of Exeter.

-Suárez E. (2008) “A room for new faces: evolution and the growth of bioinformatics,” workshop *Writing Genomics: Historiographical Challenges for New Historical Developments,* Max Planck Institute for the History of Science, Berlin.

-Suárez E. and Ramillon V. (upcoming) *Writing Genomics: Historiographical Challenges for New Historical Developments* (Max Planck Institute for the History of Science Preprint Series).

-Sutcliffe J.G. (1995) "pBR322 and the advent of rapid DNA sequencing" in *Trends in Biochemical Sciences,* 20 (2).

-Synge R. and Williams E. (1990) "Albert Charles Chibnall" in *Biographical Memoirs of Fellows of the Royal Society,* 35.

-Thurtle P. (1998) "Electrophoretic apparatus" in R. Bud (ed.) *Instruments of Science: An Historical Encyclopaedia* (London: Science Museum).

-Tutton R. (2008) "Banking expectations: the promises and problems of biobanks," International Conference of the UK Genomics Network and the Dutch Centre for Society and Genomics, Amsterdam.

-Watson J. (1969) *The Double Helix: A Personal Account of the Discovery of the Structure of DNA* (New American Library).

-Watson J. (2003) *DNA: The Secret of Life* (Heinemann).

-Watson J. and Crick F. (1953a) "Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid" in *Nature,* 171.

-Watson J. and Crick F. (1953b) "Genetical implications of the structure of deoxyribonucleic acid" in *Nature,* 171.

-Weatherall M. and Kamminga H. (1992) *Dynamic Science: Biochemistry in Cambridge, 1898-1949* (Cambridge Wellcome Unit Publications).

-Wills C. (1991) *Exons, Introns and Talking Genes: The Science behind the Human Genome, Project*.

-Wu R. (1994) "Development of the primer-extension approach: a key role in DNA sequencing" in *Trends in Biochemical Sciences,* 19.

-Wu R. and Kaiser A. (1968) "Structure and base sequence in the cohesive ends of bacteriophage lambda DNA" in *Journal of Molecular Biology,* 35, pp. 523-37.

-Wu R. and Taylor E. (1971) "Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA" in *Journal of Molecular Biology*, 57, pp, 491-511.