

# Data, Theory, and Scientific Belief in Early Molecular Biology

Pauling's and Crick's Conflicting Notions About the Genetic  
Determination of Protein Synthesis and the Solution to the  
'Secret of Life'

*Ute Deichmann*

**Abstract:** Motivated by the contradictory arguments about the relationship between data and theory in science, such as the holistic, data relativizing Duhem/Quine thesis of the underdetermination of theory by data, and a new empiricism, according to which the availability of large amounts of data is a sufficient basis for objective science, I analyze, both historically and conceptually, the generation of two highly important conflicting theories in early molecular biology: Linus Pauling's structural and Francis Crick's informational theory of biological specificity and protein synthesis. My goals are: (1) To explore the relationship between experimental data, knowledge, and theory in Pauling's and Crick's theories. (2) To show that both Pauling and Crick based their views on only a few, and almost the same, experimental data, evaluating them, however, from the different perspectives of structural chemistry and informational biology. (3) To argue that despite the apparent equivalence of data, the theories themselves were not equivalent, scientific theory choice was possible on the basis of knowledge beyond the direct experimental data, and that in general data does not speak for itself.

**Keywords:** *underdetermination, subjectivity, protein synthesis, template theories, sequence hypothesis.*

## 1. Introduction

Starting with the Human Genome Project and the availability of unprecedented amounts of sequencing data and powerful computational techniques, a new type of biology that is 'data-driven', not 'hypothesis-driven', has been proclaimed (Aebersold *et al.* 2000) that has given rise to a controversy about

the efficacy and efficiency of ‘data-driven’ versus ‘hypothesis-driven’ research (Weinberg 2010, Golub 2010). These conflicting opinions about the precedence of data or theory raise a number of questions, such as: Is there ‘data-driven’ research without prior hypothesis, and is there ‘hypothesis-driven’ or ‘theory driven’ research without relevant prior data? If hypotheses and theories are not determined by data, what are they based on?

The question of how hypotheses or theories are generated was considered philosophically irrelevant by mainstream 20th-century philosophers of science, in particular Karl Popper, but also other followers of logical empiricism and critical rationalism. The distinction between a ‘context of discovery’ and ‘context of justification’, where only the latter was considered amenable to a rational analysis, was an important argument in the justification of scientific rationality (see *e.g.* P.A. Kirschenmann 1991).

Other philosophers and scientists regarded the relationship between data and theory very important. Their answers to the questions above strongly depended on philosophical preferences that lie between the two positions of anti-empiricism and empiricism. These positions are here exemplified by the Duhem/Quine thesis of the underdetermination of theories by data on the one hand, and old and new empiricist approaches on the other. Following Larry Laudan (1990), I use the term theory in a very general way, not clearly distinguished from hypothesis, namely as any universal statement that purports to describe and explain phenomena of the natural world.

### 1.1 The underdetermination thesis

Around 1900 the physicist Pierre Duhem proposed what became known as the notion of underdetermination of theory by empirical evidence: he held that in physics the evidence available to us at a given time is insufficient to determine what scientific beliefs we should hold in response to it. His view was closely related to his holism, namely that a physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; among his examples were Newton’s particle theory and Huygens’s wave theory of light (Duhem [1906] 1954).<sup>1</sup>

While Duhem confined his views to physics, the philosopher Willard Quine, around 50 years later, extended ‘underdetermination’ to all other areas, including logic. He held that all theories are so underdetermined by the data that a scientist can, if he wishes, hold on to any theory he likes (Quine 1951). According to Larry Laudan, who critically revealed the assaults on methodology in which the underdetermination thesis was used as a justification, Quine’s holism implied a demonstration of the ambiguity of falsification: the failure of a prediction falsifies only a block of theories as a whole; it does not show which of the theories is wrong. Quine thought that only sys-

tems or complexes as a whole can be falsified, but that the choice between individual theories in them is radically underdetermined (Laudan 1990).

The thesis that any scientific theory is necessarily underdetermined by evidence has been used since the 1980s as a rationale for the claim that theory choice must to a large extent be the result of extra-scientific factors, such as social processes of ‘negotiation’ and personal interest. This claim has been well received in science studies, because it provides a general and logical reason for the necessity of social and other factors in the designation of the content of scientific theories (see *e.g.* Hesse 1981, Laudan 1990, Hacking 2000, Norton 2003).

## 1.2 Causal-mechanistic methodology and a new empiricism

In contrast to the underdetermination thesis, empiricist and positivist approaches consider sense perceptions and observational data to be a sufficient basis for the generation of theories. However, with biology becoming an experimental science at the end of the 19th century, and a molecular science in the middle of the 20th century, the generation of hypotheses, experimental testing, and a causal-mechanistic methodology became dominant epistemologies on which models and theories were based.

In recent years, powerful sequencing techniques, new computing tools, and high-throughput technology have impacted many areas of biology and provided new tools for tackling formerly largely unapproachable phenomena. The importance of data is especially large in genomics that assumed “one of the most prominent positions in terms of raw data scale across all the sciences” (Fábio *et al.* 2019). The new technologies have often been fruitfully included into causal-mechanistic research at a systems level. But they have also promoted data-driven research that does not go beyond establishing correlations, particularly in translational biomedical research. The ‘big-data revolution’ in the biomedical sciences is most visible in the rise of large institutions of translational research whose work is based on big-data technology such as the establishment of the Broad Institute on human genomics by Harvard University and MIT and of institutions of personalized medicine, such as the Center for Genomics and Personalized Medicine of Stanford University.

The ‘big data’ revolution in various sciences and also outside science has led some commentators to pronounce a ‘new era of empiricism’. It is characterized by a new epistemology which marginalizes hypotheses, experimental testing, and the search for causes, and relativizes knowledge prior to the application of an algorithm, as well as by a new scientific paradigm, namely data-intensive exploratory science (critical overviews of ‘big data’ epistemologies are *e.g.* in Kitchin 2014 and Rieder & Simon 2017). Some think that in this ‘new era of empiricism’, the large volume of “data accompanied by tech-

niques that can reveal the truth, enables data to speak for themselves free from theory” (Kitchin 2014). Though for now most biologists would disagree with such a statement, there is a tendency to conduct genomic research related to diseases with ‘big data’ alone without experimental hypothesis testing. In such data focused research, the problem of underdetermination would no longer be relevant.

This present paper aims at assessing the relationship of data, theory, and personal factors such as scientific belief in a prominent case in the history of molecular biology, namely Linus Pauling’s structural theory, and Francis Crick’s informational theory of biological specificity and protein synthesis.<sup>2</sup> This case also highlights notable differences between a chemical and a biological perspective.

I show that both Pauling and Crick based their hypothesis on experimental data, but only a few, and nearly the same ones. I claim that despite this apparent ‘underdetermination’, theory choice was possible at the time when the direct data was complemented by logic, causal analysis, and broad knowledge also outside the field in question, and that this case supports the view that experimental data are important for the generation of hypotheses or theories in biochemistry and molecular biology, but taken alone do not determine them.

## 2. Linus Pauling’s Structural Theory of Biological Specificity and Protein Synthesis

### 2.1 The theory and its background

Linus Pauling (1901-1994) was one of the greatest American chemists of the first half of the 20th century. He rebuilt chemistry on the foundation of quantum physics through his significant contributions to our understanding of chemical bonding and chemical structure. He wrote three of the century’s most influential chemistry books, among them his 1939 textbook *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* that has played a crucial role in chemical education. In the late 1920s he became a pioneer in the physical chemistry of biological macromolecules, in particular proteins. In 1951 he discovered two secondary structures, the  $\alpha$  helix and the  $\beta$  sheet, as stable conformations of polypeptide chains (his subsequent attempt to elucidate the structure of DNA was, however, not successful). In 1954 he was awarded the Nobel Prize

in Chemistry ‘for research into the nature of the chemical bond and its application to the elucidation of complex substances’.

Pauling was an early advocate of the idea that biological specificity had a molecular basis, namely the three-dimensional shape of proteins. The concept of biological specificity has played a crucial role in biology as a modern experimental science since its beginnings in the nineteenth century. At first a morphological-organismic concept related to the specificity of organisms, species, and higher taxonomic ranks, it was later conceptualized as biochemical and informational specificity. Pauling’s basic idea of specificity being based on specifically shaped molecules proved of great value to the advance of biological sciences. However, his belief that proteins’ three-dimensional structures were created by moulding proteins to the template of other proteins, antigens, or genes and that their amino acid composition or sequence were not relevant, was not fruitful and shown to be mistaken.

Pauling expressed this view since the mid-1930s. In a lecture in 1937 he considered “the secret of life itself” to lie in “how a protein molecule is able to form, out of an amorphous substrate, new protein molecules that are made after its own image” (cited in Strasser 1999). He promoted the view that templates and the complementarity of molecular structures were responsible for all biological specificity; for example, he believed that genes served as templates on which the enzymes which are responsible for the chemical characters of the organism are moulded (Strasser 2006). Pauling’s structural-chemical three-dimensional template theory became the prevailing view on protein synthesis and biological specificity and remained so until the early 1960s.

## 2.2 Data on which the theory was based

### 2.2.1 *Experiments by Pauling himself and his colleague Alfred Mirsky on reversible and non-reversible protein denaturation (i.e. the loss of their specific properties)*

The experiments gave rise to the structural theory of proteins, according to which a native protein molecule (showing specific properties) “consists of one polypeptide chain which continues without interruption throughout the molecule”, and “is folded into a uniquely defined configuration, in which it is held by hydrogen bonds” (weak bonds between two molecules) (Mirsky & Pauling 1936). Pauling and Mirsky proposed that all proteins were long chains of amino acids – a view that was not yet generally accepted at the time – that differed in their properties because of their shape, and they attributed the characteristic specific properties of native proteins “to their uniquely de-

finer configurations” (the folded shape of proteins are today called conformations).

### *2.2.2 The demonstration by immunochemists of the existence of a multitude of specific antibodies in one organism*

Since 1910, antibodies were even found against substances that had been chemically modified, *i.e.* not naturally occurring substances. These phenomena provided a strong support for ‘template theories’, in which an antigen directly modifies, or ‘instructs’, the shape of a normal serum protein to confer upon it the appropriate specificity. Unlike Paul Ehrlich’s earlier ‘selective theory’ of antibody formation, in which randomly produced pre-existing antibodies were ‘selected’ by antigens, template theories did not postulate a wasteful production of large varieties of antibodies prior to encountering an antigen, but followed the ‘rule of parsimony’, promoted by the physicist and philosopher Ernst Mach, according to which Ehrlich’s theory was ‘uneconomical’.

The immunologist Niels Jerne has pointed out that the generation of antibodies is one of several cases in the history of biology in which an ‘instructive’ theory was initially proposed to account for the underlying mechanism, but was later replaced by a ‘selective’ theory (Jerne 1967). This distinction between ‘instructive’ and selective theories was also used in the field of culture. Citing Jerne, the founder of generative linguistics Noam Chomsky considered this distinction relevant for the understanding of the generation of language in development (Jenkins 2000, p. 84).

Pauling himself put forward an ‘instructive’ template theory of antibody formation in 1940, according to which antigens ‘instruct’ the ends of globulin molecules to fold complementarily to their structure, thus becoming specific against these antigens.<sup>3</sup> He held that “all antibody molecules contain the same polypeptide chains as normal globulin, and differ from normal globulin only in the configuration” (Pauling 1940).

### *2.2.3 Experimental data from Pauling’s research on the molecular basis of sickle-cell anaemia during the 1940s and early 1950s that reinforced Pauling’s three-dimensional template theory of protein specificity and assumption of the irrelevance of amino acid sequence most strongly*

Sickle cell anaemia was discovered in 1910; its Mendelian inheritance, assumed early, was confirmed in 1949.<sup>4</sup> Pauling had studied the binding of oxygen to haemoglobin (Hb) since 1935, and when he learned that only deoxygenated blood of sickle cell anaemia patients had sickle-shaped red cells, he assumed that Hb was probably somehow involved in the sickling process. His belief in the central importance of protein shape for function led him to

hypothesize that Hb changed its shape in the sickled cells. In the mid-1940s he started a research project in order to test this hypothesis.<sup>5</sup> In 1949 his student Harvey A. Itano succeeded to clearly demonstrate a difference in electrophoretic mobility and thus electrical charge between normal and abnormal Hb (Pauling *et al.* 1949). Through chromatographic analyses that he conducted with Walter Schroeder he believed to have excluded the possibility that a difference in amino acid composition between normal Hb and sickle cell Hb caused the electrophoretic differences.

Based on these results, Pauling put forward the bold hypothesis that sickle cell anaemia was related to the structure of Hb and thus was a molecular disease caused by alterations in the shape of the Hb molecule. He was convinced that “the polypeptide chains involved in sickle-cell-anemia globin are the same as those involved in normal adult human globin”. The difference in structure between these molecules, Pauling held, was “simply a difference in the way in which the polypeptide chains are folded” (Pauling 1952). Pauling believed that normal and abnormal haemoglobin molecules were “composed of the same polypeptide chains” and that “the gene responsible for the sickle cell abnormality is one that determines the nature of the folding of polypeptide chains, rather than their composition” (cited in Strasser 2006).

#### 2.2.4 Protein chemists' claims between the 1930s and 1950s of having found numerical regularities in proteins' amino acid compositions

Despite inconclusive evidence in support of the hypotheses of numerical regularities in the amino acid compositions,<sup>6</sup> Pauling favored them (Pauling & Niemann 1939). If the hypotheses had been confirmed, the amino acid sequence would have to be excluded from being the basis of protein specificity (or to use later terminology, from carrying biological information). The importance of irregularity or aperiodicity for biological specificity was highlighted by Erwin Schrödinger in his book *What is Life* (1944), in which he likened a gene to an aperiodic crystal carrying the code of life, an idea that proved influential for Watson and Crick's work on the structure of the DNA double helix in 1953. Pauling, who as a postdoctoral fellow had spent some time in Schrödinger's laboratory studying quantum mechanics, did not mention code or aperiodicity in his later work on proteins. It should be added that some fibrous proteins, such as silk and collagen, in fact possess some kind of regularities in their amino acid sequences; thus glycine is found in almost every third position of the polypeptide chain of collagen. However, the amino acids in antibodies or in proteins that are involved in regulatory processes in the cell or in the body, in particular in enzymes and hormones, do not have a regular sequence.

Pauling propagated his three-dimensional template theory of protein synthesis until at least the late 1950s. If he referred to new contradictory exper-

imental data at all, such as Frederick Sanger's demonstration that amino acids in at least one protein, insulin, were not periodically arranged (1952) and that species differ in the amino acid sequence of insulin (1955), he integrated them into his theory as auxiliary assumptions. Thus in his proposal of a two-step protein synthesis he tried to accommodate Sanger's findings, suggesting that genes determine amino acid sequences in the first step, while proteins' shapes are determined by outside templates such as antigens (Strasser 2006).

### 3. Francis Crick's Informational Theory of Biological Specificity and Protein Synthesis

#### 3.1 The theory and its background

Francis Crick (1916-2004) was one of the most successful early molecular biologists, elucidating, together with James Watson, the three dimensional structure of DNA and, among many other things, contributing decisively to the solution of the genetic code. His numerous distinctions included the 1962 Nobel Prize in Physiology and Medicine that he was awarded together with James Watson and Maurice Wilkins 'for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material'.

Crick began his career as a physicist conducting research for a Ph.D. thesis in physics (which he never completed) at the University College London from 1937 to 1939. At the outbreak of the Second World War he began to work for the military, first designing magnetic and acoustic mines to destroy ships, and after the war working for the naval intelligence. Losing interest in working for the military and in physics, he increasingly became fascinated with basic questions of biology and the molecules of life, for example in what he called the great mystery of life, namely how the division between the living and the non-living was decided. His transition to biology was influenced by his reading of Schrödinger's *What Is Life* (1944) and Pauling's *The Nature of the Chemical Bond and the Structure of Molecules and Crystals* (1939), and he privately studied organic chemistry and biology. In 1949 he joined Max Perutz's project in Cambridge on the X-ray diffraction of proteins, and since there were no textbooks on the method yet, he learned the mathematics, needed to transform X-ray patterns into models of molecular structures, himself (Crick 1988, p. 46). As remembered by a friend, he often sought "to grasp the theory by a combination of imagery and logic, and only later to slog through the algebraic details" (Logothetis 2004). As will be shown below,



imagery and logic (in addition to scientific rigor) remained his guiding principles throughout his career. In 1954 he completed his Ph.D. thesis with Perutz on ‘X-ray diffraction: polypeptides and proteins’.

A year earlier, as is well-known, Crick completed his most acclaimed scientific work, the elucidation of the DNA double helix structure, together with James Watson, with whom he had collaborated in Cambridge since 1951 (Watson & Crick 1953a/b). For Crick, this discovery had no less importance than that they had “found the secret of life” (Watson 1968, p.115): the structure of a molecule that appeared to provide for a molecular explanation of both gene replication and biological specificity, summed up in a term used for the first time by these authors, namely ‘genetical information’ (Watson & Crick 1953b). The history of this discovery has been amply commented on including in Watson’s and Crick’s autobiographies (Watson 1968, Crick 1988; a pioneering book among the historical treatises on the subject is Olby 1974; see also Judson 1996, Morange 1998, 2020, and Sarkar 2007). Here I only highlight the paragraphs of Watson and Crick’s 1953 papers that relate to Crick’s theory of protein synthesis.

Watson and Crick made it clear that in the DNA double helix molecule, “the sequence of bases on one chain is irregular” (Watson & Crick 1953b), a statement that was reminiscent of Schrödinger’s idea of ‘aperiodic crystal’ with which he compared genes and that, unlike the periodic crystals in the inorganic world, could carry the ‘code of life’ (Schrödinger 1944). They also rendered the idea of code more concrete and introduced, for the first time, the term ‘genetic information’:

The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information. [Watson and Crick 1953b]

However, they were unable to attribute a concrete meaning to ‘code’ or ‘genetical information’. Concepts by mathematicians about code and information (see below) were not helpful in addressing concrete problems in biology (below). The prevailing theory for the genetic determination of the synthesis of specific proteins was Pauling’s template theory: genes provide physical templates for protein structures, not instructions for amino acid sequences.

In contrast, four years after his and Watson’s discovery, Crick generated a completely different theory, namely that genes provide nucleotide sequences, that are the information necessary for the determination of the amino acid sequences, and indirectly proteins’ specific shapes (Crick 1957). Crick developed these ideas in discussions with Sydney Brenner (Crick 1958).

## 3.2 Data on which Crick's theory was based

### 3.2.1 Data used by and/or known to Pauling

It is interesting that Crick based his theory largely on experimental data that Pauling either used or at least knew about, although in some crucial cases he assessed it differently. Among them were:

- Pauling's and others' work on protein structure and other elementary facts about proteins, such as their composition of only 20 amino acids, and a high probability of a given protein having a fixed composition and folding (Crick 1958).
- Experiments about function and structure of DNA on which he and Watson based their DNA double helix model. They showed that genes most probably consisted of DNA and suggested that the 1953 DNA model was correct. Pauling, too, admitted this, realizing that his own DNA model was mistaken.
- The information discourse in the sciences that started in the late 1940s. 'Information' became a key scientific concept through the work of the US National Research and Development Committee, which was set up by President Roosevelt to fund scientific research into military problems.<sup>7</sup> Among its members were the mathematician Claude Shannon, the author of *The mathematical theory of information*, 1948, and Norbert Wiener, who published *Cybernetics* in the same year. At a Symposium in California in 1948, which was attended by biologists such as Max Delbrück and George Beadle, Hungarian-American mathematician and computer scientist John von Neumann envisaged a biological equivalent of a computing machine able to reproduce itself with a set of instructions corresponding to a gene, in which he regarded the gene as an 'information tape' that could program the organism (Cobb 2013). Beadle and Delbrück were colleagues of Pauling's at the California Institute of Technology, and Pauling had written a joint paper with Delbrück (on the nature of intermolecular forces) some years earlier. But there is no indication that the 'information discourse' affected Pauling's thinking on protein synthesis.

In general, this discourse did not impact genetic or biochemical research in any concrete way. Code and information became and remained vague ways of interpreting genetic phenomena rather than precise theoretical frameworks (Kay 2000). According to Matthew Cobb (2017), Shannon's strict version of information turned out to be, for the time being, pretty much of a dead end in biology, at least regarding the concept of genetic information. Instead, as Lily Kay (2000) em-

phasized, information became an important and motivating metaphor in biological research.

It can be assumed that Crick knew of the information discourse – Watson co-authored a hoax paper on the fashionable usage of cybernetics in bacterial genetics in 1953, shortly before his and Crick’s publication of the double helix model (Cobb 2013; for the growth of ‘information’ in popular discourse, see also Cobb 2015). But there is no indication that this discourse played a motivational role in Crick’s focusing on the problem of genetic information. He was rather interested in the concrete chemistry behind this phenomenon, and the most important data for his ‘sequence hypothesis’ were provided by chemistry. In particular, there were:

- Experimental results by protein chemists questioning the periodicity hypothesis of protein structure. Crick was following most closely the work by Frederick Sanger, who in the 1950s not only showed that amino acids were not periodically arranged, but who also, for the first time, showed that their sequence was species-specific: the insulin sequences of four species of mammals were slightly different (Crick 1958). Unlike Pauling, Crick attributed high significance to Sanger’s results: “It can be argued that these [amino acid] sequences are the most delicate expression possible of the phenotype of an organism and that vast amounts of evolutionary information may be hidden away within them.” (Crick 1958) However, he had to admit that there was no evidence so far to prove that these differences of amino acid sequences between species were under the control of Mendelian genes, which was necessary to relate DNA base sequences to amino acid sequences.

### 3.2.2 Data that Crick actively sought for

Convinced that DNA controls the synthesis of proteins by determining their amino acid sequence, Crick actively sought for experimental evidence, research that is discussed in more detail here. He knew that there were no methods yet to sequence DNA and RNA. The only way to proceed was therefore to use mutants, *e.g.* of bacteria, and analyze the amino acid sequences of their proteins, something, Crick believed, had become possible at least for small proteins with the techniques developed by Sanger (Crick 1988, 102-103). In 1954, the question of a visiting geneticist, Boris Ephrussi, whether perhaps cytoplasmic genes (a vague short-lived concept) determined the amino acid sequences, whereas nuclear genes just folded proteins correctly – a suggestion that is reminiscent of Pauling’s two-step protein synthesis – convinced Crick that in order to show the centrality of sequence in nuclear

genes, he had to show that a single mutation in a nuclear gene changed one amino acid in the protein it coded for (Crick 1988, p. 103).

To this effect he started to collaborate with a newly arrived colleague at the Cavendish laboratory, the protein chemist Vernon Ingram, who had come to England as a Jewish refugee from Nazi Germany. Ingram agreed to collaborate with Crick on this genetic problem, the first time an attempt was made to bridge the two fields of genetics and biochemistry by experiments at the molecular level. Their efforts to find mutations in the enzyme lysozyme in various strains of chickens did not reveal positive results (Crick 1988, p. 105).

They started a new approach when they received some sickle cell Hb from Max Perutz, the head of the laboratory. Crick knew Pauling's work on sickle cell Hb, and he knew that human sickle cell-anaemia was the most prominent case "where a Mendelian gene has been shown unambiguously to alter a protein" (Crick 1958). He also knew that Pauling's methods were too crude to detect single changes in amino acid composition in proteins. Using Sanger's newly developed technique of fingerprinting, a combination of electrophoresis and chromatography, Ingram began to search for differences in the amino acid sequences between normal and sickle cell Hb.

In 1956 Ingram succeeded in confirming the change of one amino acid conclusively: in sickle cell Hb, one glutamic acid residue was exchanged by a valine one (Ingram 1956). This was the first demonstration of a change in amino acid sequence caused by a Mendelian gene, and Crick could state that "until recently it could have been argued that this was perhaps not due to a change in amino acid sequence, but only to a change in the folding" but that it had been "conclusively shown by my colleague, Dr Vernon Ingram" that "the gene does in fact alter the amino acid sequence" (Crick 1958).

Relating to the possible "surprise" that such a result – "the alteration of one amino acid out of a total of about 300 can produce a molecule which (when homozygous) is usually lethal before adult life" – can cause in many people, he added: "For my part, Ingram's result is just what I expected." (Crick 1958) The background for this expectation is as follows: Ingram at first thought to have found a change in two amino acids, and only when Crick urged him to repeat the experiment, he confirmed that there was only a change in one amino acid. Crick wanted and expected this result, because sickle-cell anaemia was caused by one mutation, and a change in two amino acids would have suggested an overlapping code (Crick 1988, p. 105; Olby 2009, p. 264) that had just been ruled out experimentally by Sydney Brenner (Brenner 1957, Hayes 1998). An overlapping code would mean that one DNA base was contained in more than one codon.

### 3.3 The 'sequence hypothesis'

Because this little data could be logically linked to established knowledge about the structure of proteins and DNA, Crick considered it to be a strong confirmation for his belief in the centrality of sequences and to be sufficient to suggest a concrete meaning of the term 'information': "The specification of the amino acid sequence of the protein" (Crick 1958). There was no evidence yet that amino acid sequences determined the three-dimensional shape of proteins, and most researchers believed that "the synthesis of the polypeptide chain and its folding" should be considered separately with a special mechanism leading to folding up the chain. In spite of these objections, Crick believed that "the more likely hypothesis is that the folding is simply a function of the order of the amino acids" (Crick 1958).

Realizing that a precise technique to study protein folding did not yet exist, whereas amino acid sequences could be approached experimentally, he decided to focus his reasoning on the latter, and in that sense proposed two general principles of protein synthesis, the "Sequence Hypothesis" and the "Central Dogma" (Crick 1958). For the questions of biological specificity and protein synthesis that are central to the present paper it is sufficient to look at the meaning Crick gave to his Sequence Hypothesis and not to regard the Central Dogma that has been widely commented on elsewhere (e.g. Crick 1970, Fantini 2006, Kay 2000, Morange 2006, Strasser 2006, Rosenberg 2006, Weber, 2006). The Sequence Hypothesis simply stated that "the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and this sequence is a (simple) code for the amino acid sequence of a particular protein." (Crick 1958) The basic correctness of Crick's hypothesis/principle was soon experimentally confirmed and became generally accepted. Genetic information based on sequences received a central place in many fields of biology, such as genetics, development, and evolution, a fact that prompted the author of *The Eighth Day of Creation*, Horace F. Judson, to the statement that "Crick permanently altered the logic of biology" (Judson 1996).

Later, it became clear that not all DNA sequences code for amino acid sequences, something that Crick had anticipated himself. He later wrote that he "should have said that the only way for a gene to code for an amino acid sequence of a protein is by means of its base sequence. This leaves open the possibility that parts of the base sequence can be used for other purposes, such as control mechanisms (to determine if that particular gene should be working and at what rate) or for producing RNA for purposes other than coding." (Crick 1988, pp. 108-109)

This vision that parts of the sequences can be used for control mechanisms of gene expression became concrete several years later, for example in the work by molecular embryologist and systems biologist Eric Davidson,

who devoted his life to researching the ‘regulatory genome’, *i.e.* the sequences that provide for embryonic regulation. He pointed to the “fascinating world of logic and mechanism that originates in the program of the genome. I’m not talking about encoding proteins; I’m talking about how the shape of the regulatory network modules determines its function.” (Davidson 2016) This was not the three-dimensional shape of proteins but related to a topological model of gene regulatory networks (Peter & Davidson 2015).

#### 4. The Underdetermination Thesis and the Case of Pauling’s and Crick’s Contradictory Theories

The case of Pauling and Crick shows that two central biological theories in early molecular biology were indeed underdetermined by data: the two contradictory theories of biological specificity and protein synthesis were based on few, and almost the same, direct experimental data, which were in part evaluated differently. These theories consisted of several parts. Pauling’s theory said that (i) biological specificity is based on the three-dimensional protein shape, and (ii) three-dimensional templates, not sequences, determine these shapes. Crick’s theory said that (i) biological specificity is based on base sequences that determine amino acid sequences, and (ii) amino acid sequences determine protein shapes. The first part of Pauling’s theory was complementary, not contradictory to Crick’s theory, but the second part and Crick’s theory contradicted each other.

Unlike in the Duhem/Quine ‘underdetermination’ thesis, the parts were testable independently from one another, and the fact that the theories were supported by nearly the same data did not mean that they were equivalent, *i.e.* had the same explanatory or predictive power. Despite his appreciation of Watson and Crick’s 1953 papers which emphasized the role of base sequences, Pauling’s theory lacked a causal connection of protein synthesis to DNA sequences. He also suppressed evidence that called into question several template theories of generating (not reproducing) specificities. He maintained the template theory of protein synthesis by purely speculative auxiliary hypotheses, such as that of the two-step protein synthesis mentioned above. Crick’s part two was not based on any evidence, but there was no contradicting evidence either.

Pauling’s theory of biological specificity being based on the three-dimensional shape of proteins (part one of his comprehensive theory), as well as Crick’s sequence hypothesis were fruitful and reliable and, after some modifications, were regarded true. That fruitful theories did not have to be

based on much data was clearly expressed by Crick regarding his sequence hypothesis:

“The direct evidence [for the sequence principle and central dogma] is negligible;” “I shall also argue that the main function of the genetic material is to control (not necessarily directly) the synthesis of proteins. There is *a little direct evidence* to support this, but to my mind the psychological drive behind this hypothesis is at the moment independent of such evidence.” (1958) (emphasis added) The little direct evidence was the difference of one amino acid between normal Hb and sickle-cell Hb as a genetic anomaly.

That a hard scientist like Crick sees the ‘psychological drive’ for his hypotheses as more important than direct evidence may at first sound surprising. But the question is what he might have understood by ‘psychological drive’. In my opinion this ‘drive’ was based on, first, knowledge also beyond the direct field of research, such as knowledge of the fundamental importance of proteins in the cell, the species specificity of their sequence, and the uniqueness of sequence in DNA structure. It was also based on rational analysis and vision. Thus Crick was aware of the fact that protein synthesis had to be radically different from that of other macromolecules, that it had to be highly specific, and that it was, in all probability, controlled by the genetic material of the organism (Crick 1958).<sup>8</sup>

In his autobiography, he generalized the necessity of broad knowledge for carrying out good science: “It is thus not sufficient to have a rough acquaintance with the experimental evidence, but rather a deep and critical knowledge of many different types of evidence is required, since one never knows what type of fact is likely to give the game away.” (Crick 1988, p. 141) At the same time, experimental tests were crucial to him, and Crick advised theoretical biologists to approach difficulties not by tinkering with their theory but by seeking some crucial test (Crick 1988, p. 141) (something he had done himself by asking Ingram to sequence the two types of Hb). Often, it is not the amount of data that matters, but the quality of a particular piece of evidence.

The cases of Pauling’s structure theory and Crick’s information theory of biological specificity show that in the life sciences, different answers to one and the same basic question can be complementary and do not have to be contradictory (though there are also cases of rejected knowledge, such as Pauling’s theory of proteins being synthesized on three-dimensional templates). Unlike Aristotle’s principle of ‘form’ – organisms’ definition or essence – that he considered an immaterial principle, ‘information’ in modern biology cannot be separated from matter, *i.e.* from chemistry. It does not exist without the molecules of DNA or proteins, and chemical changes in them affect their information. Both Pauling’s structure and Crick’s information theory were materialistic theories with Crick’s theory pointing to a special relationship of different macromolecules as the basis of life.

Crick's sequence hypothesis became the basis of what is called 'big data' biology today, *i.e.* biological and medical research that utilizes the availability of large amounts of sequencing and expression data (allowing simultaneous characterization of huge amounts of an organism's genes and proteins and their interactions) and new sophisticated computational techniques. But as his homage to Pauling shows, he himself valued highly not only information biology related to nucleic acids but also chemistry and the molecular approach. Contrasting Pauling with physicist-turned molecular biologist Max Delbrück who did not care much for chemistry and regarded it, in Crick's words, "as a rather trivial application of quantum mechanics", Crick himself emphasized the importance of Pauling's belief that the "well-established ideas of chemistry and, in particular, the chemistry of macromolecules" were crucial for biology. (Delbrück later realized that his underestimation of chemistry and biochemistry was a mistake and acknowledged that molecular biology was not a trivial aspect of biological systems.) Pauling believed, Crick wrote, "that our knowledge of the various kinds of atoms and the bonds that hold atoms together [...] would be enough to crack the mysteries of life." (Crick 1988, pp. 60-61) According to Crick, molecular biology was "at the heart of the matter", and "almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself" (Crick 1988, pp. 60-61).

The importance of molecules for understanding life has been highlighted by molecular biologist and philosopher Michel Morange who emphasized that the secret of life resides "in the systemic relationships that the chemical components of organisms jointly support": "The network of chemical relations that characterizes life on earth could not exist without a certain type of molecule. The living world is therefore the product, both structurally and functionally, of a particular chemistry." (Morange 2008, p. 143) According to Morange (2020), the greatest biological achievements of the past few decades, which also include programs such as the human genome sequencing, should still be understood within the molecular paradigm.

## 6. Conclusion

(i) The Duhem/Quine thesis of underdetermination of theory by data and its implication for the impossibility of choosing between different theories based on the same data, does not apply to Pauling's and Crick's theories of biological specificity and protein synthesis. It is true that these theories, like many theories or models in the history of biology, were based on only a few and nearly the same experimental data, and that they (the theories) were con-



tradictory on one of their two central aspects. But unlike what was proposed in the Duhem/Quine thesis (a) the two parts of the theories could be tested separately; (b) the theories were not equivalent, and theory choice was possible at the time on the basis of logic and knowledge outside the field in question; (c) in their second main aspect, the theories, though different, were complementary and not mutually exclusive.

(ii) Though data were not sufficient, they were important for the generation of fruitful hypotheses.

Pauling's and Crick's theories, like most hypotheses and theories in the history of experimental biology, were not determined by experimental data alone, but data were very important. Both, Pauling and Crick actively sought to find experimental evidence that would support their hypotheses/theories. Only when they found some evidence that they thought was crucial did they propose their hypotheses.

(iii) 'Underdetermination' does not render the generation of hypotheses and theories arbitrary or purely psychological, as the proponents of the distinction of a context of discovery and context of justification suggest.

The fact that the generation of hypotheses is not only based on direct experimental evidence does not render this process arbitrary. In the cases of Pauling and Crick, what was behind the generation of a new hypothesis, in addition to new direct evidence, was a broad knowledge of the field in question and of other related fields, and logical considerations, something that has often been called intuition.

The integration of approaches from different fields that were previously quite separate played a major role in the generation of Crick's 'sequence hypothesis' or 'principle': Crick's and Ingram's collaboration on the molecular basis of a genetic disease was the first bridge at the molecular level between the formerly separate fields of genetics and chemistry. This collaboration resulted in providing evidence that, small as it was – a change of one amino acid out of a total of around 300 – Crick regarded as sufficient to support his bold hypothesis.

(iv) The necessity of subjectivity for good science. Pauling's theory of biological specificity based on protein shapes and spatial complementarity, and Crick's sequence principle, had a long-lasting impact on biology, where they were experimentally confirmed in their very essence right to the present time. Crick acknowledged the important role of subjectivity in good science, when he wrote, "I had always appreciated that the scientific way of life, like the religious one, needed a high degree of dedication and that one could not be dedicated to anything unless one believed in it passionately." (Crick 1988, p. 17) The examples of Crick and Pauling support physicist-philosopher Michael Polanyi's dictum that good science is not only based on data and exper-

imental testing, but also on knowledge, competence, and a ‘passionate commitment’ to ‘a vision of reality’ (Polanyi 1959).

Computer scientist Judea Pearl has critically looked at the tendency in mainstream statistics to always grant data priority over opinions and interpretations, because data is deemed objective. According to Pearl, “unlike correlation and most of the other tools of mainstream statistics, causal analysis requires the user to make a subjective commitment [...] Where causation is concerned, a grain of wise subjectivity tells us more about the real world than any amount of objectivity.” (Pearl & Mackenzie 2018, pp. 88-89) Pearl believes that “some statisticians to this day find it extremely hard to understand why some knowledge lies outside the province of statistics and why data alone cannot make up for lack of scientific knowledge.” (*Ibid.*) Knowledge and causal analyses are cornerstones in experimental biology, where the concept of genomic causality is today at the core of genetics, developmental biology, and evolution.

As the case of Pauling shows, the inclusion of subjective decisions and beliefs into scientific theories runs the risk of leading an unwary researcher astray. This risk, however, is usually reduced in a critical scientific community. However, as Pauling’s and others’ adherence to template theories of antibody generation shows, even when they became increasingly questionable or obsolete, incorrect and plainly wrong beliefs in fashionable fields promoted by prominent scientists sometimes have a long life (see also Deichmann & Müller-Hill 1998, Deichmann 2007). A reliance on data alone, however, would have prevented the generation of most of the seminal theories and concepts in the history of modern biology, among them the generation of Pauling’s concept of molecular specificity and molecular disease, as well as Crick’s ‘sequence hypothesis’ (sequence principle) that became the basis of the use of large amounts of sequencing data in genomics, proteomics, and other fields of ‘big data’ biology today.

## Acknowledgements

I thank Anthony S. Travis and Klodian Coko and two anonymous referees for valuable comments on an earlier draft of the paper.

## Notes

<sup>1</sup> According to some interpretations of Duhem, the underdetermination thesis follows from his holistic testing. The proposition of a hypothesis in physics required

auxiliary assumptions. If experiments falsify the hypothesis, you don't only falsify the hypothesis but also the auxiliary hypotheses involved. Different scientists will try to accommodate the discrepancies in different manners, and thus underdetermination arises (Coko 2015, chapter 5; and Zammito 2004, chapter 2).

- <sup>2</sup> Bruno Strasser (2006) has pointed to the two competing theories of protein synthesis from a different perspective than in this article.
- <sup>3</sup> Pauling relates here to the idea of complementary structures for antibody and antigen that was suggested by Friedrich Breinl and Felix Haurowitz some ten years earlier.
- <sup>4</sup> James V. Neel (1949) showed that sickle cell anaemia was a manifestation of a homozygous condition, whereas sickle cell anemia was the result of a heterozygous condition.
- <sup>5</sup> Bruno Strasser (1999) has examined the details of Pauling's sickle cell anaemia research and its social background.
- <sup>6</sup> A close observer of this research at the time, protein chemist Joseph Fruton, held that in these alleged discoveries "chemists and biochemists have been led astray by their penchant for numerology because they allowed imagination and intuition to take precedence over attention to the limits of their experimental procedures and the validity of their empirical data." (Fruton 1972, pp. 151-2)
- <sup>7</sup> This account of the influence of the information concept into the sciences is based on Kay (2000) and Cobb (2017).
- <sup>8</sup> Crick referred to biochemist Alexander Dounce who pointed to the dilemma that protein synthesis cannot be brought about by enzyme catalyzed processes with the enzymes being synthesized by other protein-enzymes and so forth, and who had started "to look for possible solutions that would be acceptable, at least from the standpoint of logic. The dilemma, of course, involved the specificity of the protein molecule, which doubtless depends to a considerable degree on the sequence of amino acids in the peptide chains of the protein. The problem is to find a reasonably simple mechanism that could account for specific sequences without demanding the presence of an ever-increasing number of new specific enzymes for the synthesis of each new protein molecule." (Crick 1958) Crick concluded that "the synthesis of proteins must be radically different from the synthesis of polysaccharides, lipids, co-enzymes and other small molecules; that it must be relatively simple, and to a considerable extent uniform throughout Nature; that it must be highly specific, making few mistakes; and that in all probability it must be controlled at not too many removes by the genetic material of the organism." (Crick 1958)

## References

- Aebersold, R.; Hood, L.E. & Watts, J.D.: 2000, 'Equipping Scientists for the New Biology', *Nature Biotechnology*, **18**, 359.
- Brenner, S.: 1957, 'On the Impossibility of all Overlapping Triplet Codes in Information Transfer from Nucleic Acid to Proteins', *Proceedings of the National Academy of Sciences of the United States of America*, **43**, 687-694.
- Cobb, M.: 2013, '1953: When Genes Became "Information"', *Cell*, **153**, 503-506.

- Cobb, M.: 2015, *Life's Greatest Secret: The Story of the Race to Crack the Genetic Code*, London: Profile Books.
- Cobb, M.: 2017, '60 Years Ago, Francis Crick Changed the Logic of Biology', *PLoS Biology*, 2017, 1-8, (online available at <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.2003243&type=printable>, accessed 1 January 2020).
- Coko, K.: 2015, *The Structure and Epistemic Import of Empirical Multiple Determination in Scientific Practice*, Dissertation Thesis, Bloomington: Indiana University.
- Crick, F.H.C.: 1957, 'Protein Synthesis', Lecture at the Society for Experimental Biology Symposium on the Biological Replication of Macromolecules, University College London, 19 September 1957, cited in Cobb (2017).
- Crick, F.H.C.: 1958, 'On Protein Synthesis', *Symposia of the Society for Experimental Biology*, 12, 138-163.
- Crick, F.H.C.: 1970, 'Central Dogma of Molecular Biology', *Nature*, 227, 561-563.
- Crick, F.H.C.: 1988, *What Mad Pursuit: A Personal View of Scientific Discovery*, New York: Basic Books.
- Davidson, E.H.: 2016, 'Interview', Caltech, 14 December 2013 by U. Deichmann, *Developmental Biology* 412, S20-S29.
- Deichmann, U. & Müller-Hill, B.: 1998, 'The Fraud of Abderhalden's Enzymes', *Nature*, 393, 109-111.
- Deichmann, U.: 2007, "'Molecular" versus "Colloidal": Controversies in Biology and Biochemistry, 1900-1940', *Bulletin for the History of Chemistry*, 32, 105-118.
- Duhem, P.: [1906] 1954, 'Physical Theory and Experiment', in: P. Duhem, *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press, pp. 183-190.
- Fábio, C.P.; Navarro, H.M.; Chengfei, Y.; Shantao, L.; Mengting, G.; Meyerson, W. & Gerstein, M.: 2019, 'Genomics and Data Science: an Application within an Umbrella', *Genome Biol.*, 20, 109.
- Fantini, B.: 2006, 'Of Arrows and Flows; Causality, Determination, and Specificity in the Central Dogma of Molecular Biology', *History and Philosophy of the Life Sciences*, 28, 567-593.
- Fruton, J.S.: 1972, *Molecules and Life: Historical Essays on the Interplay of Chemistry and Biology*, New York: Wiley-Interscience.
- Golub, T.: 2010, 'Counterpoint: Data first', *Nature*, 464, 679.
- Hayes, B.: 1998, 'The Invention of the Genetic Code', *American Scientist*, 86, 8-14.
- Hacking, I.: 2000, *The Social Construction of What?* Cambridge, MA: Harvard University Press.
- Hesse, M.: 1981, 'Revolutions and Reconstructions in the Philosophy of Science', *Philosophy*, 56, 430-431.
- Ingram V.M.: 1956, 'A Specific Chemical Difference between the Globins of Normal Human and Sickle-cell Anemia Haemoglobin', *Nature*, 178, 792-794.
- Jenkins, L.: 2000, *Biolinguistics. Exploring the Biology of Language*, Cambridge: Cambridge University Press.
- Jerne, N. K.: 1967, 'Antibodies and Learning: Selection versus Instruction', in: G.C. Quarton *et al.* (eds.), *Neurosciences*, New York: Rockefeller University Press, pp. 200-205.
- Judson, H. F.: 1996, *The Eighth Day of Creation: Makers of the Revolution in Biology*, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
- Kay, L.E.: 1993, *The Molecular Vision of Life*, Oxford: Oxford University Press.
- Kay, L.E.: 2000, *Who Wrote the Book of Life? A History of the Genetic Code*, Cambridge: Cambridge University Press.

- Kirschenmann P.A.: 1991, 'Local and Normative Rationality of Science: the Content of Discovery Rehabilitated', *Journal for General Philosophy of Science*, **22**, 61-72.
- Kitchin, R.: 2014, 'Big Data, Epistemologies and Paradigm Shifts', *Big Data & Society*, **1**, 1-12.
- Laudan, L.: 1990, 'Demystifying Underdetermination', in: C.W Savage (ed.), *Minnesota Studies in the Philosophy of Science*, **14**, 267-297.
- Logothetis, N.K.: 2004, 'Francis Crick 1916-2004', *Nature Neuroscience*, **7**, 1027-1028.
- Mirsky, A.E. & Pauling, L.: 1936, 'On the Structure of Native, Denatured, and Coagulated Proteins', *Proceedings of the National Academy of Sciences*, **22**(7), 439-447.
- Morange, M.: 1998, *A History of Molecular Biology*, Cambridge, MA: Harvard University Press.
- Morange, M.: 2006, 'Protein Side of the Central Dogma: Permanence and Change', *History and Philosophy of the Life Sciences*, **28**, 513-524.
- Morange, M.: 2008, *Life Explained*, New Haven: Yale University Press.
- Morange, M.: 2020, *The Black Box of Biology. A History of the Molecular Revolution*, Cambridge: Harvard University Press.
- Neel, J.V.: 1949, 'The Inheritance of Sickle Cell Anemia', *Science*, **110**, 64-66.
- Norton, J.D.: 2003, 'Must Evidence Underdetermine Theory?', In: M. Carrier, D. Howard & J. Kourany (eds.), *The Challenge of the Social and the Pressure of Practice*, Notre Dame-Bielefeld Interdisciplinary Conference on Science and Values; Zentrum für Interdisziplinäre Forschung, Universität Bielefeld, July 9-12 [online available at [https://www.pitt.edu/~jdnorton/teaching/1702\\_jnrsnr\\_sem/1702\\_jnrsnr\\_seminar\\_2005/docs/underdet\\_thesis.pdf](https://www.pitt.edu/~jdnorton/teaching/1702_jnrsnr_sem/1702_jnrsnr_seminar_2005/docs/underdet_thesis.pdf), accessed 13/1/2020].
- Olby, R.: 1974, *The Path to the Double Helix: The Discovery of DNA*, New York: Dover Publications.
- Olby, R.: 2009, *Francis Crick. Hunter of Life's Secrets*, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Pauling, L. & Niemann, C.: 1939, 'The Structure of Proteins', *Journal of the American Chemical Society*, **61**, 1860-1867.
- Pauling, L.: 1939, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, Ithaca: Cornell University Press.
- Pauling, L.: 1940, 'A Theory of the Structure and Process of Formation of Antibodies', *Journal of the American Chemical Society*, **62**, 2643-2657.
- Pauling, L.; Itano, H.A.; Singer S.J. & Wells I.C.: 1949, 'Sickle Cell Anemia, a Molecular Disease', *Science*, **110**, 543-548.
- Pearl, J.P. & Mackenzie, D.: 2018, *The Book of Why: The New Science of Cause and Effect*, New York: Basic Books.
- Peter, I. & Davidson, E.H.: 2015, *Genomic control process: Development and evolution*, Amsterdam: Academic Press.
- Polanyi, M.: [1958] 1962, *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago: University of Chicago Press.
- Quine, W.V.O.: 1951, 'Two Dogmas of Empiricism', *The Philosophical Review*, **60**, 20-43.
- Rieder, G. & Simon, J.: 2017, 'Big Data: A New Empiricism and its Epistemic and Socio-Political Consequences', in Pietsch, W. Wernecke, J., Ott, M. (eds.), *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big*

- Data*, pp. 85-105, Berlin: Springer [available online: [https://link.springer.com/chapter/10.1007/978-3-658-12153-2\\_4](https://link.springer.com/chapter/10.1007/978-3-658-12153-2_4), accessed 1 August 2020].
- Rosenberg, A.: 2006, 'Is Epigenetic Inheritance a Counterexample to the Central Dogma?', *History and Philosophy of the Life Sciences*, **28**, 549-565.
- Sarkar, S.: 2007, *Molecular Models of Life: Philosophical Papers on Molecular Biology*, Cambridge, MA: The MIT Press.
- Schrödinger, E.: 1944, *What is Life?* Cambridge: Cambridge University Press.
- Strasser, B.: 1999, 'Sickle Cell Anemia, a Molecular Disease', *Science*, **286**, 1488-1490.
- Strasser, B.: 2006, 'A World in One Dimension: Linus Pauling, Francis Crick and the Central Dogma of Molecular Biology', *History and Philosophy of the Life Sciences*, **28**, 491-512.
- Watson, J.D.: 1968, *The Double Helix. A Personal Account of the Discovery of the Structure of DNA*, New York: Simon and Schuster.
- Watson, J.D. & Crick, F.H.C.: 1953a, 'A Structure of Deoxyribonucleic Acid', *Nature*, **171**, 737-738.
- Watson, J.D. & Crick, F.H.C.: 1953b, 'Genetical Implications of the Structure of Deoxyribonucleic Acids', *Nature*, **171**, 964-967.
- Weber, M.: 2006, 'The Central Dogma as a Thesis of Causal Specificity', *History and Philosophy of the Life Sciences*, **28**, 595-610.
- Weinberg, R.: 2010, 'Hypotheses First', *Nature*, **464**, 678.
- Zammito, J.H.: 2004, *A Nice Derangement of Epistemes; Post-Positivism in the Study of Science from Quine to Latour*, Chicago: Chicago University Press.

*Ute Deichmann:*

*Jacques Loeb Centre for the History and Philosophy of the Life Sciences,  
Ben-Gurion University of the Negev, Beer Sheva 8496500, Israel;  
uted@post.bgu.ac.il*