Studies in History and Philosophy of Biological and Biomedical Sciences

Stud. Hist. Phil. Biol. & Biomed. Sci. 33 (2002) 1–28

www.elsevier.com/locate/shpsc

Strategies in the interfield discovery of the mechanism of protein synthesis

Lindley Darden [a], Carl Craver [b]

[a] *Department of Philosophy, University of Maryland, College Park, MD 20742, USA*
[b] *Department of Philosophy, Florida International University, 3000 N.E. 151st Street, North Miami, FL 33181, USA*

Received 25 July 2000; received in revised form 21 March 2001

Abstract

In the 1950s and 1960s, an interfield interaction between molecular biologists and biochemists integrated important discoveries about the mechanism of protein synthesis. This extended discovery episode reveals two general reasoning strategies for eliminating gaps in descriptions of the productive continuity of mechanisms: schema instantiation and forward chaining/backtracking. Schema instantiation involves filling roles in an overall framework for the mechanism. Forward chaining and backtracking eliminate gaps using knowledge about types of entities and their activities. Attention to mechanisms highlights salient features of this historical episode while providing general reasoning strategies for mechanism discovery.

Keywords: Mechanism; Protein synthesis; Strategies; Interfield relations; Discovery; Molecular biology; Biochemistry

> A spectacular display of progress in knowledge of the mechanism of protein synthesis has taken place during the past decade . . .
>
> (Zamecnik, 1969, p. 1)

E-mail addresses: darden@carnap.umd.edu (L. Darden); craverc@fiu.edu (C. Craver).

1369-8486/02/$ - see front matter
PII: S1369-8486(01)00021-8

> What is found in biology is *mechanisms*, mechanisms built with chemical components . . .
>
> (Crick, 1988, p. 138)

1. Introduction

Crucial pieces of the mechanism of protein synthesis were discovered by biochemists and molecular biologists in the 1950s and 1960s. At the outset, the approaches of these fields were very different, focusing on different components and finding different aspects of the mechanism from different ends. By about 1965, the results from the different approaches were integrated. The scientific work leading to this integration reveals general strategies for discovering mechanisms.

This instance of interfield integration, like many other discovery episodes in the biological sciences, crucially involves discovering a mechanism. Focusing centrally on *mechanisms* provides new ways of thinking about discovery, interfield integration, and reasoning strategies for scientific change. Philosophers of science have separately analyzed scientific discovery, interfield relations, mechanisms and reasoning strategies. This paper brings together these disparate topics. A unified approach yields reasoning strategies in discovering mechanisms that integrate results from different fields.

Many philosophers of science (for example, Popper, 1965) have been skeptical about finding methods for reasoning in discovery. Even those who have had much to say about scientific change (Kuhn, 1962; Laudan, 1977; Kitcher, 1993) have not so much as discussed reasoning strategies for discovering new paradigms, traditions or practices. Nonetheless, a few philosophers have worked on discovery (for example, Nickles, 1980a,b; Meheus & Nickles, 1999). Reasoning in discovery is a more tractable problem if discovery is viewed as an extended process of construction, evaluation and revision (Darden, 1991). One fruitful reasoning strategy is to search for a solution to a problem in one field by relating it to items in another field (Darden & Maull, 1977; Bechtel 1984, 1986; Darden, 1991). A new perspective is emerging by focusing on interfield relations in the discovery of mechanisms.

Some philosophers have argued for the importance of mechanisms in science (Wimsatt, 1972; Brandon, 1985; Glennan, 1996; Machamer, Darden & Craver, 2000) and in molecular biology in particular (Burian, 1996; Crick, 1988). Wimsatt, for example, says that, ‘At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms . . .’ (Wimsatt, 1972, p. 67). In their pioneering work, Bechtel and Richardson (1993) elucidate decomposition and localization strategies for discovering mechanisms in simple and complex systems.

This paper discusses two additional strategies for interfield mechanism discovery: schema instantiation and forward chaining/backtracking. Schema instantiation is the application of an abstract mechanism framework and the search for components to fill in its details. The strategy of forward chaining/backtracking involves reasoning

about known (or hypothesized) mechanism components to fill gaps in the understanding of the productive continuity of the mechanism, either forward or backward.

The protein synthesis case is a rich case for investigating reasoning in an interfield discovery episode. Molecular biologists and biochemists brought different ideas and techniques to the problem of how proteins are synthesized. At the outset, they worked on different ends of the mechanism. Eventually, their results were integrated to produce a single description of the mechanism. Protein synthesis comprises one of the core mechanisms in the two fields: the mechanism for gene expression in molecular biology, and the mechanism for the synthesis of enzymes and structural proteins important in the study of metabolism in biochemistry. Consequently, understanding this interfield discovery episode illuminates reasoning in a significant achievement of twentieth-century biology that integrated results from two fields, molecular biology and biochemistry.

Although they are in need of reinterpretation from the perspective of reasoning in interfield mechanism discovery, valuable historical accounts of parts of the protein synthesis story exist. The history of this episode has been told both by the scientists involved, in autobiographical accounts (for example, Crick, 1988; Gros, 1979; Hoagland 1990, 1996; Jacob, 1988; Watson, 1962, 1968, 2000; Zamecnik 1962a,b, 1969), and by historians of molecular biology and biochemistry (for example, Burian, 1996; Chadarevian & Gaudillière 1996, 1996; Gaudillière, 1993; Judson, 1996; Kay, 2000; Morange, 1998; Olby, 1970; Thieffry & Burian, 1996). There are aspects of this story (mostly its biochemical side) that we will not discuss. However, we will weave together some of the new historical work by Rheinberger (1997) concerning Paul Zamecnik's biochemical research with more familiar accounts of molecular biological work. Rheinberger's emphasis on Zamecnik's 'experimental system' highlights a previously neglected aspect of experimental scientific work, as well as the changes in the problem-contexts in which that system was used. However, Rheinberger neglects Zamecnik's explicit discussion (for example, Zamecnik, 1969) of the use of that experimental system in the search for the *mechanism of protein synthesis*. Unlike previous work, we stress interfield integration through, and strategies for, mechanism discovery.

We begin with a brief characterization of mechanisms, the abstract schemata that are used in their description and their discovery, and two reasoning strategies that make use of this characterization, namely schema instantiation and forward chaining/backtracking. In Section 3 we interpret the discovery of protein synthesis as a case of interfield mechanism discovery, paying particular attention to the strategies used in this scientific episode. The final section abstracts from examples in the case study, and discusses the two general strategies for mechanism discovery in more detail.

2. Mechanisms, schemata, strategies

An abstract characterization of mechanisms aids in analyzing this historical case and in finding strategies for mechanism discovery: 'Mechanisms are entities and

activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (Machamer et al., 2000, p. 3). Types of entities include ions, macromolecules (such as proteins and the nucleic acids, DNA and RNA) and cellular structures, such as ribosomal particles, which are composed of both RNA and proteins. Types of activities include geometrico-mechanical activities, such as lock and key docking of an enzyme and its substrate, and electrochemical activities, such as strong covalent bonding and weak hydrogen bonding.

Entities having certain kinds of properties are necessary for the possibility of acting in certain ways, and certain kinds of activities are only possible when there are entities having certain kinds of properties. Entities and activities are interdependent (Machamer et al., 2000, p. 6). For example, appropriate chemical valences are necessary for covalent bonding, polar charges are necessary for hydrogen bonding, and appropriate shapes are necessary for lock and key docking.

Mechanisms are made of components that work together to do something. The entities and activities are organized in productive continuity from beginning to end. One goal in discovering a mechanism is to reveal the mechanism's productive continuity. Determining the temporal boundaries of the mechanism, that is, the set-up and the termination conditions, allows work to proceed from both ends to find the intermediate stages. Looking forward, each stage must give rise to, allow, drive or make the next. Conversely, looking back, each stage must have been produced, driven or allowed by the previous stage(s). Consequently, the reasoning strategy of forward chaining from the (perhaps hypothesized) set-up conditions and backtracking from the termination conditions is a fruitful research strategy for finding the productive continuity of a mechanism.

In addition to using components in the forward/backward strategy, another strategy for discovering a mechanism is schema instantiation. Mechanism schemata are abstract frameworks for mechanisms. They contain place-holders for the components of the mechanism (both entities and activities) and indicate, with variable degrees of abstraction, how the components are organized. Often these place-holders characterize a component's role in the mechanism. Discovering a mechanism involves specifying and filling in the details of a schema, that is, instantiating it by moving to a lower degree of abstraction. As we will see, diagrams and equations are often employed to depict graphically the schematic organization of mechanisms.

3. Discovering the mechanism of protein synthesis 1953–1965: biochemistry and molecular biology

The discovery of the mechanism of protein synthesis was an interfield discovery. Both biochemists and molecular biologists contributed to it.

A group of biochemists were working to understand a mechanism for assembling polypeptides. They took the end of the mechanism to be a protein, consisting of amino acids held together by strong covalent bonds. By the 1940s, when MD-turned-biochemist Paul Zamecnik began his work, biochemists had discovered over twenty amino acids and had elucidated the nature of the linkages between them in peptide

bonds. Zamecnik and his colleagues, especially Mahlon Hoagland, sought to understand energetic intermediates between free amino acids and their linkage in polypeptides (recalled in Zamecnik, 1962a, 1969; Hoagland, 1990, 1996). They were thus working backward from peptide bonds to the mechanisms of polypeptide assembly, focusing on chemical reactions and energy requirements for such strong covalent bonds to form. Biochemists often used *in vitro* experimental systems, such as Zamecnik's cell-free rat liver preparation. As Zamecnik put it graphically: 'The biochemist traditionally studies living cells by smashing them to bits and trying to analyze the function of their parts' (Zamecnik, 1958, p. 118).

The other players in the episode were molecular biologists, such as James Watson and Francis Crick, who took the beginning of the mechanism to be DNA. They sought to understand the 'genetic code', as it came to be called, by which the order of the bases in DNA is related to the order of amino acids in proteins. They were thus reasoning forward from DNA to ordered amino acid sequences in proteins. Molecular biologists focused on weak hydrogen bonds and on determining macromolecular structure. Their experimental techniques were often grounded in X-ray crystallography and the building of scale models, which had earned a good reputation in Watson and Crick's work on the structure of DNA (Watson & Crick, 1953a) and Linus Pauling's work on protein structure (Pauling & Corey, 1950). Molecular biologists also used genetic techniques, such as crossbreeding, to investigate the role of DNA in genetic mechanisms.

Zamecnik contrasted the approaches of the two fields:

> As in the building of a tunnel, digging is going on from two sides of this mound of uncertainty, in the hope of meeting in the middle. Investigators primarily interested in protein synthesis are moving back to a study of ribonucleic acid metabolism, while those interested primarily in the gene and DNA metabolism are studying interrelationships with ribonucleic acid from the other side. It has become quite clear that ribonucleic acid is the connecting link between the hereditary message of the gene and its enzymic expression (Zamecnik, 1962a, p. 47).

These molecular biologists and biochemists differed from each other in the techniques they used, in the parts of the mechanism they investigated, and in their attention to different aspects of the productive continuity in the mechanism. While biochemists, such as Zamecnik, were homogenizing rat livers and tracing centrifuge fractions, some molecular biologists were crystallizing macromolecules and subjecting them to X-ray analysis, and yet others were doing genetic crosses with bacteria and the bacterial virus, bacteriophage. While biochemists were subjecting the protein end of the mechanism to chemical analysis, molecular biologists began with the genetic material of DNA. While molecular biologists traced the 'flow of information' (Crick's phrase: Crick, 1958, p. 144), the biochemists studied the flow of matter or energy in the mechanism. While molecular biologists questioned how the genetic information contained in the order of bases along the DNA double helix served to order the amino acids in proteins, biochemists investigated the energy

requirements for binding free amino acids in the strong, covalent, peptide bonds of proteins.

Despite these numerous differences, fruitful interfield interactions between molecular biology and biochemistry served to integrate their findings. As work proceeded from both ends of the mechanism, it converged in the middle, as we will see, with the discovery of new entities (three types of RNA and activated amino acids) and their activities. The result was an understanding of the productive continuity of the protein synthesis mechanism from beginning to end.[1]

3.1. Diagrams of 1953–4

The differences between these fields are nicely illustrated by comparing Zamecnik's and Watson's diagrams in Fig. 1.

Zamecnik's diagram of 1953 shows the components of an in vitro, cell-free protein synthesis system. As Rheinberger (1997) ably recounts, Zamecnik's experimental system was constructed from homogenized and centrifuged components of rat liver cells. Work with this system provided Zamecnik with a mechanism sketch illustrated in the diagram. It includes several components (some of which are numbered in the diagram): continued energy production by means of the formation of ATP (a molecule which supplies energy) from precursors (1); radioactive amino acids that were added to the system by the researchers (2); and a centrifuge fraction (unnumbered) coming in from below, which was presumed to contain enzymes. These all come together on the microsomes (3). Zamecnik later showed that the microsome had as a component a ribonucleoprotein particle, later still called a 'ribosome', which was the functional unit (Zamecnik, 1958). The microsomes were hypothesized to be the location of protein synthesis, as the arrow pointing to 'incorporation' indicates. 'Incorporation' traces the flow of radioactive counts from the labeled amino acids to what was presumed to be (and later shown to be) a polypeptide chain. This chain consisted of amino acid subcomponents that were covalently bonded with peptide bonds. For Zamecnik and his colleagues, the termination condition to be understood was peptide bond formation, and understanding this required finding its energy requirements, as well as the possible intermediates between free amino acids and polypeptide chains (Zamecnik, 1969).

Zamecnik's lab group concentrated their investigations on the stages of the mech-

[1] We take this to be a successful case of discovery in science. It is difficult to imagine that proteins are not usually synthesized by the mechanism involving DNA and RNA that we discuss here, even though, since the 1960s, much more has been discovered about this mechanism. Furthermore, some special case anomalies are now known, such as in RNA viruses and in cases where amino acids are added or removed post-translationally. The discovery of introns, RNA splicing and RNA editing added additional stages to the mechanism in eucaryotes. We have neglected the history of work on the enzymes that catalyze various stages in this mechanism; an exciting new chapter begins with the recent hypothesis that the ribosome itself functions as a ribozyme in the formation of the peptide bond (Cech, 2000); we thank Jeff Lewandowski for calling this to our attention. However, even if the proposed mechanism schema of protein synthesis is revised in the light of subsequent evidence, the reasoning strategies discussed here are still general ones for discovering plausible mechanisms.

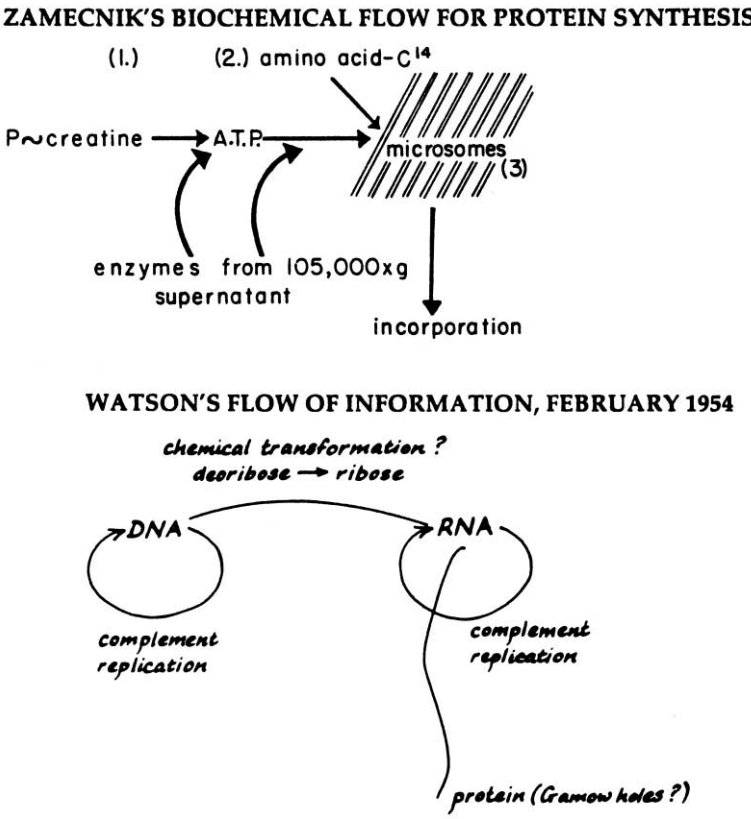

Fig. 1. Biochemical and molecular biological sketches for protein synthesis (from Judson, 1996, p. 273; used with permission of Cold Spring Harbor Press, James Watson and Paul Zamecnik).

anism near the end product. This continued a tradition in biochemistry, begun earlier in the century, in which the nature of the peptide bond and the chemical formulae of amino acids had been elucidated (Zamecnik, 1969; Rheinberger, 1997). Biochemical and cytological studies had shown that RNA was part of the microsomes and was somehow associated with protein synthesis. (Brachet's and Caspersson's work on this is discussed in Thieffry & Burian, 1996; Rheinberger, 1997.) However, it was unclear in 1954 what the RNA did. It was not even a schematic place-holder in a biochemical equation; it filled no biochemical role.

In contrast, Watson's 1954 diagram (see the lower part of Figure 1) follows the flow of genetic information from DNA to protein via an RNA intermediate stage. The line labeled 'protein (Gamow holes?)' refers to an hypothesized geometrico-mechanical activity for the RNA in determining amino acid order. Watson's idea was that DNA was copied into RNA, which would form a structure with differently shaped holes. Then, different amino acids would fit into the holes. The RNA template would thus determine the order in which the amino acids bonded in the protein.

RNA did play a role in the molecular biological schema of DNA→RNA→protein. It filled a gap in the flow of information between DNA and protein.

Watson was altering an idea proposed by the physicist George Gamow (1954). Gamow had proposed that proteins were synthesized directly on the DNA double helix, by fitting into 'holes' in the helix (discussed in detail in Kay, 2000, Ch. 4). The holes were the spaces between the turns of the helix. Gamow thought that different groups of three or four bases in the DNA might produce differently shaped holes. This was his way of filling in the brief suggestion of Watson and Crick (1953b) that it 'seems likely that the precise sequence of the bases is the code which carries the genetical information'.

Gamow speculated about the nature of the code, that is, the relation between a particular group of bases (presumed to be 3 or 4) and a particular amino acid (discussed in Kay, 2000). However, Gamow was unaware of the biochemical and cytological evidence that proteins are not synthesized directly on DNA but are associated with RNA in the cytoplasm. Consequently, the entity whose structure was relevant was at a subsequent stage of the mechanism; it was not the structure of DNA but of RNA. Carrying Gamow's idea forward to the next stage, Watson proposed that the cytoplasmic RNA had 'holes', whose shapes were determined by the RNA bases surrounding the holes. Different amino acids were assumed to fit into different holes, bringing them into close spatial proximity so that peptide bonds could form. Finally, the protein would leave the RNA template. In 1954, Watson speculated that RNA might have a helical shape (Rich & Watson, 1954a,b; discussed in Watson, 1962, 2000).

The contrasts between Zamecnik's and Watson's schematic mechanism diagrams graphically depict the differences between the approaches of biochemists and molecular biologists mentioned briefly above. They had different starting points. Watson began with DNA, Zamecnik with the energy requirements of peptide bond formation. The molecular biologists wished to show how the order of the bases in the DNA (the genetic material) determined the order of the amino acids in the protein. Zamecnik's lab investigated high-energy intermediates back along the path to free amino acids.

The differences between the fields were clearly recognized at the time. Hoagland, a colleague in Zamecnik's lab, drew this contrast in recounting his 1955 announcement of amino acid activation at a meeting of molecular biologists:

> The palpable indifference with which the audience received the news showed how tightly closed the door between the biochemistry and molecular biology compartments was. The focus of the meeting, of course, was on how the polymerizing system ordered its units, not on how it energized them (Hoagland, 1996, p. 78).

Zamecnik also reflected back on the differences:

> If molecular biology is the domain of large molecules, the study of protein synthesis was rooted in the simple biochemical desire to understand how the energy barrier from free amino acid to peptide was overcome. The numerous connections

of protein synthesis with the nucleotide-nucleic acid family were unanticipated in the early days (Zamecnik, 1984, p. 466).

Despite these differences, researchers in both fields recognized the importance of understanding the mechanism of protein synthesis. Both were speculating about some role for RNA. They followed each other's work, even though the molecular biologists had little interest in the energy requirements of the mechanism. Zamecnik recounted a visit by Watson in 1954 during which they discussed the newly discovered double helix model of DNA. Zamecnik expressed disappointment that the bases are inside the helix; it seemed as if they were less available to play some role in protein synthesis (Zamecnik, 1969, p. 5). Researchers in both fields knew of Frederick Sanger's work in the late 1940s and early 1950s in which he determined the sequence of amino acids in insulin and thereby showed that various prior hypotheses about protein structures were incorrect (recalled in Crick, 1988; Zamecnik, 1962a; discussed in Chadarevian, 1996). They approached the role of RNA from both ends, with their own techniques and perspectives on the mechanism, moving forward from DNA and backward from peptide bond formation.

3.2. Molecular biological work: three-dimensional structure

In 1954, Watson and Rich attempted to find the structure of RNA by using the same techniques that had led to the discovery of the double helical structure of DNA. They produced fibers and took X-ray photographs, but the results were inconclusive and did not provide sufficient evidence for the hypothesized RNA helix (Rich & Watson, 1954a,b).

RNA viruses, such as tobacco mosaic virus (TMV), produce their own proteins in the host. In 1952, Watson had learned X-ray crystallography techniques by investigating the structure of TMV (Watson, 1968). TMV is rod-shaped, but other small RNA viruses are spherical. Crick and Watson (1957, p. 12) noted that spherical RNA viruses have a similar RNA content, as well as a similar shape, to the microsomes of cells. In their work on viral structure, Crick and Watson speculated that the amino acid sequence of the viral protein 'is determined by the molecular structure of the RNA of the infecting virus' and 'the "coding" implied . . . is relatively simple' (Crick & Watson, 1957, p. 7). However, their hypotheses of the structure of viruses (Crick & Watson, 1956) did not illuminate the structure of the microsomes or the code for relating RNA bases to amino acids. RNA viruses were not sufficiently like microsomes to provide a simpler system for studying the RNA intermediate in protein synthesis more generally.

In sum, the X-ray techniques, the studies of viral structure and the search for a presumed three-dimensional structure of RNA did not bring the same success for the structures and activities of RNA in protein synthesis as it had for the structure of DNA and the mechanism of DNA replication. Extending the research program from the success with DNA structure was a plausible approach, but these early molecular biological efforts did not elucidate the mechanism of protein synthesis.

3.3. Draining the biochemical bog

While Watson and Crick were investigating RNA structure, Zamecnik's lab was busy 'draining the biochemical bog' (Rheinberger, 1997) of the homogenized rat liver system for cell-free protein synthesis. They were grinding up rat livers and subjecting them to centrifugation in order to separate components and thereby to investigate their relative roles in the mechanism. Such work allowed the discovery of previously unknown entities in centrifuge fractions that were required for the mechanism to operate. By 1954, the cell-free system for the incorporation of ^{14}C-labeled amino acids into a polypeptide chain was working. In addition to the amino acids, the system required microsomes, an undifferentiated centrifuge fraction (the 105,000×g supernatant) and ATP (a molecule that provides energy), but no DNA. When Hoagland joined Zamecnik's lab, his research focused on the question of whether an intermediate step occurred in the reaction, between free amino acids and the formation of peptide bonds (Zamecnik, 1969; Hoagland, 1990). This question was within the traditional biochemical approach of finding intermediates in chemical reactions and in investigating the energy requirements for the formation of covalent bonds, in this case the peptide bond. Hoagland did indeed find a high-energy intermediate in the reaction, as he backtracked along the pathway to free amino acids. In the biochemical notation of Zamecnik (1984), this reaction is characterized:

$$aa_1 + pppA + E_1 \rightleftharpoons aa.pA_1 \cdot E_1 + pp$$

The amino acid (aa) combines with ATP (adenosine triphosphate, here pppA), and the reaction is catalyzed by an enzyme (E). Two phosphates (pp) are released and the other product of the reaction is an activated amino acid, called 'aminoacyl-AMP' ($aa.pA_1$). This reaction provided 'a mechanism for activation of amino acids' (Hoagland, 1955, p. 288). Thus, as Hoagland backtracked from bound amino acids in proteins to free amino acids, he found the first of two intermediates. This aminoacyl-AMP was the expected high-energy intermediate of an activated amino acid. The other was a bit of RNA that changed from being 'junk' to having an important role in the mechanism, a role not anticipated in the biochemical reaction schemata (Hoagland, 1990, Ch. 5; Rheinberger, 1997). Transfer RNA was one of several RNAs that would be found to play roles in the mechanism and whose discovery would fill gaps in the understanding of its productive continuity.

3.4. Adaptor RNA and soluble RNA to transfer RNA

After the biochemical work had shown that the ribonucleoprotein particles, not the lipoprotein membranes of microsomes, are associated with protein synthesis (Zamecnik, 1958), Watson began work on three-dimensional ribosomal structure. He thus continued to pursue his view that the three-dimensional structure was the crucial property of RNA for enabling its activity (discussed in Watson, 1962, 2000).

In contrast to the reaction equations of the biochemists, the molecular biological notation depicts information flow. As of about 1957, the sequential stages in the

mechanism schema might have been depicted in the following way: one DNA sequence→one nucleoprotein particle in the microsome→one protein chain.

Crick began to think about the requirements to build a three-dimensional structural model of RNA with twenty differently shaped holes. As he recounted:

> Well, since we can't find the structure of RNA from the data, let's do it the other way around by assuming that there are twenty different cavities and trying to build a structure that *had* twenty different cavities. And as soon as you put it that way, you saw that it was almost impossible to *do* (quoted in Judson, 1996, p. 283).

Some of the amino acids have very similar shapes. Unable to construct any possible scale model, Crick became skeptical of Watson's 'holes in RNA' hypothesis.

Crick hypothesized using the weak electro-chemical activity of hydrogen bonding, not the geometrico-mechanical activity of Watson's schema, for the role of ordering amino acids. Crick recounts his reasoning:

> The main idea was that it was very difficult to consider how DNA or RNA, in any conceivable form, could provide a direct template for the side-chains of the twenty standard amino acids. What any structure *was* likely to have was a specific pattern of atomic groups that could form hydrogen bonds. I therefore proposed a theory in which there were twenty adaptors (one for each amino acid), together with twenty special enzymes. Each enzyme would join one particular amino acid to its own special adaptor. This combination would then diffuse to the RNA template. An adaptor molecule could fit in only those places on the nucleic acid template where it could form the necessary hydrogen bonds to hold it in place. Sitting there, it would have carried its amino acid to just the right place it was needed (Crick, 1988, pp. 95–6).

This idea was first suggested in 1955 in a widely circulated but unpublished paper (Crick, 1988, p. 95). In its first published form, Crick speculated about the nature of the molecules that played the role of adaptors: 'These might be any sort of small molecule — amino sugars, for example — but an obvious class would consist of molecules based on di-, tri-, or tetranucleotides' (Crick, 1957, p. 26). These nucleotides would be particularly suited to play the role of the complementarily charged items needed for hydrogen bonding. Nucleic acid bases have polar structures with slight charges. Such charges were already known to hold the double helix of DNA together; maybe they also played a role in the protein synthesis mechanism.

Meanwhile, Zamecnik and Hoagland had been investigating RNA synthesis as an adjunct research project to their investigation of protein synthesis. In their centrifuged preparations, they found a soluble RNA fraction that differed from the heavier microsomal RNA. To their surprise, this soluble RNA was covalently bound to the ^{14}C-labeled amino acids (Zamecnik, 1969, p. 6; Hoagland, 1990, p. 94; Judson, 1996, p. 324; Rheinberger, 1997, p. 155).

In 1956, the biochemical work in Zamecnik's lab was integrated with the molecular biology work after Watson visited Hoagland in Zamecnik's lab. Anything serving

the role of an adaptor in the protein synthesis mechanism would have to bind to amino acids. When Watson saw Hoagland's results, he told Hoagland of Crick's as then unpublished idea of an adaptor RNA that attached to the amino acids (Hoagland, 1990, p. 93).

Hoagland recalls:

> I was bowled over by the ingenuity and beauty of Francis's idea and sensed that it had to be the explanation of our experimental findings. An image arose in my mind: we biochemical explorers were hacking our way through a dense jungle to discover a beautiful long-lost temple, while Francis Crick, flying gracefully overhead on gossamer wings of theory, waited for us to see the goal he already was gazing down upon (Hoagland, 1990, p. 94).

He also told less happy stories about this episode (quoted in Rheinberger, 1997, p. 157). Naturally, scientists prefer to figure out for themselves what role in a mechanism is played by an entity they discover. Of this episode, Rheinberger, who stresses the differences between the two approaches, said: 'Biochemical reasoning in terms of metabolic intermediates came to be confronted with reasoning in terms of genetic information transfer' (Rheinberger, 1997, p. 158).

In 1959, Hoagland, Zamecnik and their colleague Mary Stephenson published a figure (see Fig. 2) that incorporated this idea of transfer RNA, the small saw-toothed pieces, which were still referred to as soluble RNA. They are shown binding to amino acids (aa) and bringing them to the supposedly helix-shaped ribosomal RNA comprising the ribosome. The saw-toothed patterns, although depicted geometrico-mechanically, were said to represent hydrogen bonding between complementary charges on bases of the soluble RNAs and the ribosomal RNA (Hoagland et al., 1959, p. 111). The solid lines along the string of amino acids represent the usual peptide bonds. This diagram integrates the perspectives of both fields. Molecular biology provided the hypothesized helical structure of the template RNA and the activity of hydrogen bonding. Biochemistry provided the activated amino acids that become covalently bonded, first to transfer RNA and then to each other in the protein.

Furthermore, the biochemical results served to revise details of Crick's adaptor hypothesis. Transfer RNA, as soluble, adaptor RNA came to be called, turned out to be larger than Crick proposed. Hoagland (1996, p. 79) said that Crick should have expected the adaptor to be larger than the coding trinucleotide because it also needed a specific active site to attach to its specific amino acid. Crick had not sufficiently considered the side reaction in the mechanism that Hoagland had extensively investigated. Prior to the coding stage (on which Crick focused), the amino acid attached to the adaptor RNA. Thus, the RNA had to be large enough to have two active sites, one to bind the amino acid, and another to hydrogen bond to template RNA. Consideration of all its roles in the mechanism would have shown that the adaptor RNA had to be larger than a di- or trinucleotide.

Again, the molecular biologists and the biochemists had concentrated on different subcomponents of the mechanism, even different activities of a given subcomponent. They arrived there by different routes. Crick was reasoning forward about the activi-

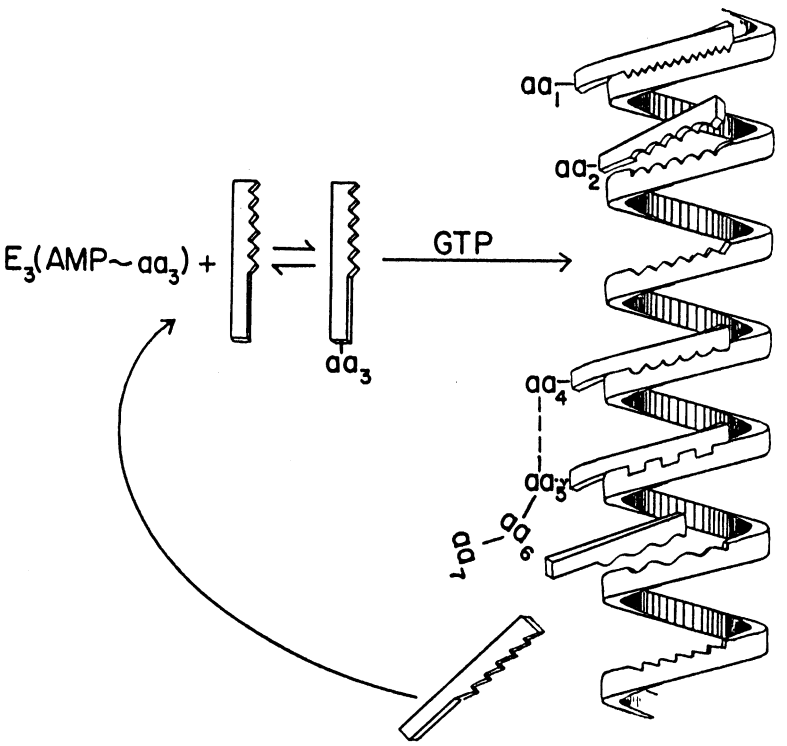

Fig. 2. A schema for the interaction of microsomal RNA and soluble RNA-amino acid (from Hoagland, Zamecnik & Stephenson, 1959, p. 110; used with permission of University of Chicago Press).

ties of nucleic acids in information transfer, while Hoagland was backtracking along the path to the activated amino acids and empirically isolating them in an *in vitro* system where he, surprisingly, found an attached RNA.

Crick gave the term 'information' a precise characterization after he stated what he dubbed the 'Central Dogma', 'that once "information" has passed into protein *it cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acids in residues in the protein' (Crick, 1958, p. 153).

3.5. Anomalies for the ribosome as template

In the late 1950s, anomalies began to emerge for the view of ribosomes as carrying the information for ordering the amino acids. In 1958, two Russians, Belozersky and Spirin, showed that 'the DNA of different microorganisms had greatly different base ratios . . . The base composition of the total RNA of these same organisms hardly varied at all . . . ' (Crick, 1959, pp. 35–6). Most of the RNA in the cell is found in

the ribosomes. If DNA is transcribed into ribosomal RNA, one would expect the base ratios of the DNA and the ribosomal RNA in a given species to be similar. They were not. This anomalous data, confirmed by two groups, challenged the role proposed for the ribosome as a template in the mechanism of protein synthesis. The ribosomal base anomaly indicated a problem about the DNA to RNA step in the proposed mechanism schema because the expected relationship between the two was not found.

Crick (1959) generated a set of alternative hypotheses to resolve this anomaly, localized in various stages of the mechanism schema. We will not take the time here to consider each of them; notably, he did not include the postulation of an as yet undiscovered type of RNA having a base composition like DNA. This is the idea of a separate messenger RNA, different from either transfer RNA or ribosomal RNA. It was by the discovery of such a messenger RNA that the ribosomal base anomaly was soon resolved in 1961. Reflecting back on this reasoning later, Crick (1988) says that there were plenty of RNAs available, for example, the ribosome had been shown to have two RNA components of different sizes. So, there seemed no need to postulate an as-yet-undetected type of RNA to serve as the expected template.

It is instructive to note the difference here between adaptor RNA and messenger RNA from the perspective of the mechanism's activities. When only one type of RNA was known, namely the ribosome, Crick needed a second type, the adaptor, to postulate hydrogen bonding. Hydrogen bonding occurs between complementary charges on different molecules, so the use of the activity of hydrogen bonding demanded a second type of molecule, probably RNA, to bond to ribosomal RNA. Hence, Crick hypothesized a second type, adaptor RNA, to play this role. No such demand was present that required messenger RNA; Crick was not forced to postulate it by reasoning forward about the activities of the mechanism or their correlative entities. Instead, it was discovered during the resolution of an empirical anomaly about the rate of protein synthesis in *in vivo* genetic experiments. (For a general discussion of anomaly driven theory change, see Darden 1991, 1995; Darden & Cook, 1994.)

3.6. Rate anomaly and messenger RNA

In addition to the ribosomal base anomaly, another empirical anomaly arose for the idea of a stable ribosome as the template for protein synthesis in bacteria. Work by Arthur Pardee, François Jacob and Jacques Monod produced the puzzling experimental result. Jacob and Monod were working on the control of protein synthesis using genetic techniques and observing the effects of mutants (Judson, 1996; Morange, 1998). In the process of studying its control mechanism, they discovered a new component of the primary mechanism of protein synthesis. As Crick said in reflecting on reasoning in mechanism discovery: one must sort out effects due to the nature of a mechanism itself from effects due to its control when trying to unscramble a complex biological system (Crick, 1988, p. 111). Forward chaining or backtracking requires sorting out the components of a primary mechanism from the components of mechanisms that control it.

In the famous, so-called PaJaMo bacterial mating experiment, protein synthesis began very quickly after a functional gene entered recipient *E. coli* bacteria lacking that gene (Pardee, Jacob & Monod, 1959; discussed in Olby, 1970; Schaffner, 1974; Judson, 1996; Morange, 1998). Ribosomes were assumed to be stable particles, requiring some time for synthesis. If the ribosomal RNA had to be synthesized on the incoming DNA (the functional gene), and then the ribosomal particle had to be assembled, one would not expect the very rapid initiation of protein synthesis. A greater time lag would be expected. Again, the molecular biologists were chaining forward from the gene to the next stage in protein synthesis; they were considering the time that the assembly of the entities (presumed to be ribosomal templates) in the next stage would require.

Three alternative hypotheses were devised to account for the surprising rapidity of the initiation of protein synthesis (Olby, 1970). Each stage in the proposed molecular biological mechanism schema served as a site of hypothesis formation. Reasoning forward from the DNA, the first hypothesis proposed that the DNA of the gene itself could serve as the site of protein synthesis. But this hypothesis was problematic because protein synthesis was already known to be associated with the ribosomes, not with DNA. Maybe protein synthesis in bacteria differed from the systems previously studied by biochemists, such as Zamecnik's cell-free rat liver system. Bacterial DNA is not in a separate nucleus but is in the cytoplasm itself. Monica Riley, Pardee's student, thought that the hypothesis of synthesis on the bacterial DNA might be the correct hypothesis (Riley, personal communication).

A second hypothesis was that ribosomes were rapidly synthesized after the DNA entered the recipient cell. But ribosomes were known to be stable particles with at least two RNA subunits, so there seemed to be insufficient time to synthesize new ones before protein synthesis started. The third hypothesis was that a new RNA, with a base sequence like that of the DNA, was synthesized quickly. This DNA-like RNA (also called X or 'tape' RNA) would then use the existing stable ribosomes as the sites for protein synthesis. (For further discussion of these three alternatives, see Olby, 1970; Jacob, 1988.)

Thus, we see that reasoning forward about the mechanism's stages and the time needed for each stage introduces temporal constraints on the description of the mechanism (Craver & Darden, 2001). If an hypothesized stage would be expected to take longer than a time-course experiment shows that it does, then the resulting anomaly indicates the need for a change in the hypothesized mechanism. Something else was needed to instantiate the role of the template, something that can form more rapidly than ribosomes, something with a base sequence like that of DNA.

Some empirical evidence already existed for a DNA-like RNA, but it had been misinterpreted (discussed in Watson, 1962; Judson, 1996, pp. 414–5, 418–22). New experiments were done to try to provide more direct evidence for this 'tape RNA' (Jacob, 1988) or 'messenger RNA', as Jacob and Monod named it in 1961:

> The property attributed to the structural messenger of being an unstable intermediate is one of the most specific and novel implications of this scheme . . . This

> leads to *a new concept of the mechanism of information transfer*, where the protein synthesizing centers (ribosomes) play the role of non-specific constituents which can synthesize different proteins, according to specific instructions which they receive from the genes through M-RNA (Jacob & Monod, 1961, p. 353; emphasis added).

This unstable intermediate in the mechanism was a new type of RNA. Messenger RNA was transcribed from DNA and thus carried the genetic code for ordering the amino acids during protein synthesis. This new type of RNA would be formed quickly, as was required by the rate at which synthesis began after the DNA entered the recipient bacterium. Flow of information, namely the order of bases and the order of amino acids, was again the focus of the molecular biologists. (Compare Crick's statement that he would not discuss the flow of energy or matter but would the flow of information in protein synthesis, in Crick, 1958, p. 144.)

New experimental work provided good evidence for the existence of messenger RNA. Sydney Brenner and François Jacob succeeded in differentially labeling old ribosomal RNA and newly synthesized messenger RNA (Brenner, Jacob & Meselson, 1961). Also, in Watson's lab, François Gros added radioactively labeled uracil to *E. coli* and detected an unstable RNA that was neither ribosomal nor transfer RNA (Gros, Hiatt, Gilbert, Kurland, Risebrough & Watson, 1961). All the different RNA components of the mechanism had been found. (For more on the discovery of messenger RNA, see Gaudillière, 1996; Gros, 1979; Jacob, 1988; Judson, 1996; Morange, 1998; Olby, 1970; Rheinberger, 1997; Watson, 1962.)

Subsequent work, utilizing both biochemical and, to a lesser extent, genetic techniques, led to cracking the genetic code (Nirenberg and Matthaei, 1961; Crick, Barnett, Brenner & Watts-Tobin, 1961). The work on the code is one with more interfield competition between biochemists and molecular biologists than the cooperation discussed thus far. That story has been told elsewhere (Crick, 1988; Judson, 1996; Kay, 2000).

From that competition emerged an integrated account of the triple, non-overlapping code operating within the protein synthesis mechanism. The base sequence in DNA is transcribed into messenger RNA, which moves to the ribosomes, the site for the subsequent stages. A specific triplet codon on the messenger RNA hydrogen bonds to its complementary anticodon on a transfer RNA, which is attached to its specific activated amino acid. As the transfer RNAs bond sequentially to the messenger RNA, the amino acids are brought into appropriate proximity so that peptide bonds form. Incorporation of amino acids occurs in a specific linear order, based on the order of the codons in the messenger RNA. The ribosomes are merely the site where the messenger RNA attaches and is held in an appropriate orientation for these stages in the mechanism to occur. The ribosome no longer played the role of template in the information flow schema; that role was now filled by the tape-like messenger RNA. The type of mechanism had changed from a mold-like template to a tape along which the ribosome moved, sequentially incorporating amino acids (Crick, 1988).

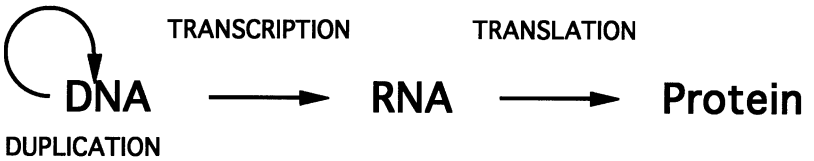

Fig. 3. Abstract schema for the central dogma (redrawn, based on Watson, 1965, p. 298).

3.7. Summary of protein synthesis case

Diagrams by Watson represent the integrated understanding of the mechanism of protein synthesis as of the 1960s. The diagrams come from Watson's *The Molecular Biology of the Gene* (Watson, 1965), the first molecular biology textbook. Just as the Zamecnik group began incorporating the molecular biological views of structure and hydrogen bonding in their 1959 diagram, Watson's more detailed 1965 diagrams integrate the biochemical findings of the activation of amino acids and their covalent bonding to soluble (later 'transfer') RNA.

Watson's abstract schema for the mechanism of protein synthesis is 'often called the central dogma' (Watson, 1965, p. 297). Fig. 3 illustrates the components of the mechanism in a very schematic way. DNA is transcribed into messenger RNA, which is translated into protein. Parts of this mechanism are depicted in much more detail in two other figures (Figs. 4 and 5). Fig. 4 shows the roles of all three types of RNAs and the other entities in the mechanism, including the DNA, studied by molecular biologists, and the amino acids, studied by biochemists. Activities include *transcription* of DNA to messenger RNA and *translation* of messenger RNA, utiliz-

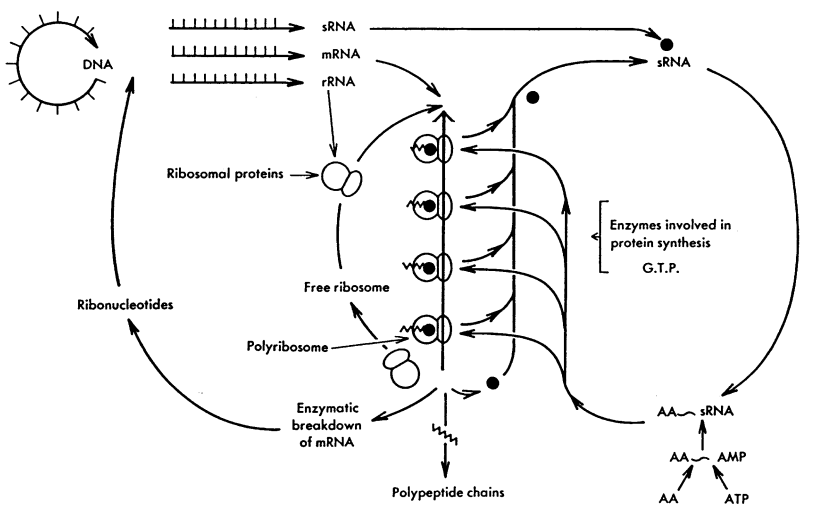

Fig. 4. 'Schematic view of the role of RNA in protein synthesis' (Fig. 11-10 in Watson, 1965, p. 338; copyright 1965 by J. D. Watson; reprinted by permission of Addison Wesley Longman, Inc.).

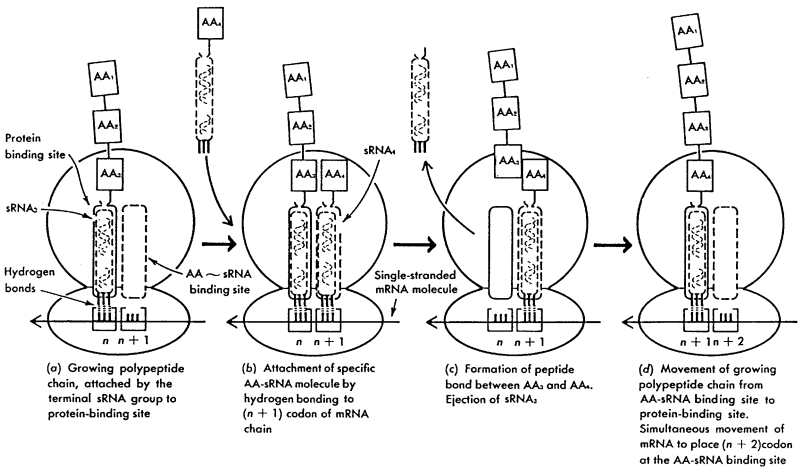

Fig. 5. 'Diagrammatic representation of the stepwise growth of a polypeptide chain' (Fig. 11-9 in Watson, 1965, p. 337; copyright 1965 by J. D. Watson; reprinted by permission of Addison Wesley Longman, Inc.).

ing transfer RNAs and ribosomes, into protein. Translation is carried out, most importantly, by the activity of *weak hydrogen bonding* between slightly charged complementary bases in the transfer RNA and the messenger RNA. The biochemical findings are shown in the lower right: amino acids (AA) combine with ATP to form the high energy intermediate (AA~AMP), which then covalently bonds to soluble RNA (sRNA), which then moves to the ribosome. Further biochemical entities are depicted: 'Enzymes involved in protein synthesis', and another energy molecule (GTP), whose role in the mechanism was then unknown. Another of Watson's figures graphically depicts the hydrogen bonding in its dotted lines. Hydrogen bonding between three bases and their complements is how the genetic code is 'read'. The ribosome, with its two circular parts, moves along the messenger RNA in the successive stages of the mechanism. The growing peptide chain of amino acids extends from the top of the diagram. Note the structural details, and the way the activity of hydrogen bonding is depicted with dots extending between the lines representing the charged bases of the transfer RNA and messenger RNA. Structures of macromolecules and weak hydrogen bonding are important features of the molecular biologist's schema of protein synthesis.

The discovery of all the key components in the mechanism of protein synthesis required the integration of aspects of the mechanism studied by both fields. Molecular biologists worked forward from the DNA. Biochemistry worked backward from peptide bonds to activated amino acids. They met in the middle and both contributed to the elucidation of the roles of the three types of RNA and the activities of hydrogen bonding, covalent bonding and energetic intermediates.

4. Strategies for discovering mechanisms

The discovery of the mechanism of protein synthesis illustrates a number of strategies for discovering mechanisms that move beyond those discussed in previous work (Darden, 1991; Bechtel & Richardson, 1993). This section discusses those strategies in a more general way. Discovery cannot be relegated to woolly flights of imagination, unconstrained conjecture, or single 'a-ha' moments. Rather, there are more or less reliable, yet inherently fallible, strategies for discovery. Discovery is construed broadly to include construction, evaluation and anomaly resolution.

4.1. Productive continuity

Mechanisms, and in particular the mechanism of protein synthesis, are entities and activities organized such that they are productive of regular changes (Machamer et al., 2000). The entities and activities are organized spatially and temporally in a productive continuity in which one stage produces or gives rise to the next (Craver & Darden, 2001). This productive continuity may be traced by following what flows or what is passed, transmitted, preserved, transformed, prevented or allowed from one stage of the mechanism to the next.

In reasoning in the discovery of mechanisms, one proceeds with the goal of eliminating gaps in the description of the mechanism's productive continuity. These gaps may appear as spatial or temporal gaps, where no role can as yet be schematized. Alternatively, a gap may simply involve an inability to specify how a given role in the mechanism schema is to be filled. In searching for the productive continuity of the mechanism, one must find an activity for each entity and an entity for each activity.

We have seen two different conceptions of this productive continuity, an informational conception proposed by molecular biologists, and a chemical/energetic conception investigated by some biochemists.

Crick delineated flow of matter, flow of energy and flow of information in the mechanism of protein synthesis and clearly focused on information flow (Crick, 1958, p. 144). What mattered to molecular biologists such as Watson and Crick was tracing information flow, that is, the preservation of linear order and pattern from one stage of the mechanism to the next. Productive continuity in this case is measured in terms of constancy (or at least partial preservation) of linear order from DNA bases through messenger RNA bases to amino acids in the protein.

For biochemists, in contrast, the productive continuity of the mechanism depends on the flow of matter and energy through various stages of the mechanism, usually depicted in chemical equations. Zamecnik sought high-energy intermediates (such as the activated amino acids) to fill the productive gap between free amino acids and strong covalent bonds holding them together in the polypeptide chains, that is, to understand how the energy barrier from free amino acid to peptide was overcome.

The chemical/energetic and the informational senses of productive continuity were integrated with the discovery of the continuous mechanism involving the flow of matter, energy and information, all together.

This case reveals two general types of strategies for reasoning to eliminate gaps

in the understanding of the productive continuity of a mechanism or, more positively, for discovering the organization that sustains that productive continuity. These strategies are, first, schema instantiation and, second, forward chaining and backtracking. We consider these in turn.

4.2. Schema instantiation

One strategy for constructing an hypothetical mechanism is to instantiate an abstract schema (Darden & Cain, 1989; Machamer et al., 2000). Such a schema provides the framework of the mechanism. Schemata typically specify roles, black boxes, at varying degrees of abstraction and with more or less detail specified. The schema terms can then be filled with the mechanism's entities and activities as they are hypothesized and discovered.

In the discovery of the mechanism of protein synthesis, biochemists and molecular biologists were using different schemata. Biochemists tried to instantiate chemical reaction schemata in chemical equations to show how an energetically unfavorable reaction could occur. These chemical schemata guided researchers to seek high-energy intermediates and led Hoagland and Zamecnik to see the importance of activated amino acids. The schemata represented in biochemists' equations exhibit the productive continuity explicitly as balance of matter, reshuffling molecules on either side of the equations. Such chemical reaction schemata suggest finding balanced equations, high-energy intermediates, enzymes and conditions leading the reaction to go in one direction rather than the other.

Molecular biologists instead used a schema depicting the flow of information, which Crick characterized as the precise determination of sequence of either bases or amino acids (Crick, 1958). That flow was further specified in Watson's version of the central dogma. (For the discussion of the contrast between Watson's and Crick's versions of the central dogma, see Keyes, 1999a,b.) Watson worked to find the pieces to instantiate the schema DNA→RNA→protein.

As of 1952, what Watson knew about the RNA portion in the middle was very sketchy, a mere black box (role), specified only as a template. Subsequent work explored different ways of filling that black box, of finding the RNA machinery, and of thereby eliminating this gap in the mechanism's productive continuity and information flow. In contrast, the biochemists' schemata had no role for RNA. Thus, when it was empirically found to be associated with protein synthesis, that produced a puzzle as to what role it played.

As discussed in Section 3, the mechanism schema of protein synthesis underwent several changes. Initially, there were two schemata for the different fields. The molecular biologists' schema,

(1m) DNA→helical RNA→protein,

ordered amino acids via a geometrico-mechanical activity.
The biochemists' schema,

(1b) amino acids+ATP+other centrifuge fractions→activated aa complex→protein,

focused instead on the energy requirements of the reaction.
These schemata were integrated in an interim schema:

(2) DNA→ribosomal RNA+activated aa–tRNA complex→protein,

with ordering via hydrogen bonding activity.
In the face of empirical anomalies for the idea of the ribosomal template and the subsequent discovery of messenger RNA, the integrated schema was then transformed to:

(3) DNA→messenger RNA+activated aa–tRNA complex+ribosomal RNA→protein,

with ordering via hydrogen bonding activity between triplet codons.

These changes show that the schema became more complicated, with one type of RNA being changed to two, then to three types. Also, the activity of hydrogen bonding came to play a crucial role in the linear ordering of the amino acids. The biochemical work on the role of ATP in the activation of amino acids was integrated into the schema, as a source of energy for peptide bond formation. These interfield relations led to the integration of previously separate schemata into a single productively continuous description of the mechanism.

Schema instantiation, however, is but one of the strategies for discovering mechanisms. Gaps in the understanding of a mechanism's productive continuity can also be filled by relying on what we know about entities, activities and their interdependencies. This knowledge is deployed in the strategies of forward chaining and backtracking.

4.3. Forward chaining and backtracking to find productive continuity

When reasoning about productive continuity in mechanisms, one can reason from either end of the mechanism, that is, by forward chaining or by backtracking. Because the choice of the beginning and the end of the mechanism is somewhat relative and context-dependent, this forward–backward strategy applies equally well to any convenient starting or ending point in a mechanism. Furthermore, this reasoning strategy can be applied to branching mechanisms, which can be followed forward or backward along side branches, and to feedback and other cyclic mechanisms, which can be arbitrarily viewed as beginning somewhere in the cycle.

Forward chaining and backtracking are made possible by the intelligible relationships between interdependent entities and activities. Entities and a specific subset of their properties enable the activities in which they engage (given appropriate conditions). Furthermore, activities require distinct types of entities and properties of those entities as the basis for such acts.

Scientists come to mechanism construction with prior knowledge about types of entities and activities that facilitates forward chaining and backtracking. Typically, a field will have a store of knowledge about types of entities and activities that figure in proposed mechanisms in that field at a given time (Craver & Darden, 2001). As we have seen, during the 1950s and 1960s in molecular biology that store included three-dimensional macromolecular nucleic acids and proteins (often having helical shapes), as well as weak types of chemical bonding, such as hydrogen bonding. Biochemists of that period (as exemplified in Zamecnik's work) typically hypothesized chemical reactions by considering smaller molecules, such as amino acids, with particular bonding properties, such as valence. Bonding activities were typically strong ones, such as covalent and ionic bonding. High energy molecules, such as ATP, and catalysts, such as enzymes, were in the biochemists' store. (In our discussion of the historical case, we omitted the discovery of the enzymes involved in the protein synthesis mechanism; that's another part of the story.)

Forward chaining and backtracking each have two subtypes, one for entities and one for activities. Attending to entities during forward chaining, one may use what is known or conjectured about the *activity enabling properties of entities*. Such attention allows one to speculate as to the kinds of activities that a given entity can engage in to produce the next stage of the mechanism. Alternatively, one may use knowledge of the occurrence of an activity in the mechanism to conjecture as to the consequences of that activity for entities and properties in the next stage. This is conjecturing about *activity consequences*.

Conversely, when backtracking, the properties of an entity can provide clues as to the activities that produced it. That is, one or more of an entity's properties may serve as an *activity signature*, a property that signals to the researcher the prior occurrence of some activity. Alternatively, during backtracking, one may find *entity signatures* of activities, that is, properties of activities that provide clues as to what entities in a prior stage may have led to the occurrence of those activities.

Consider each of these in turn.

4.3.1. Activity-enabling properties of entities

Quite general properties of entities enable the activities in which they can engage. Such properties include three-dimensional structure and size, as well as orientation and location *in situ*. Structures can promote or prohibit geometrico-mechanical activities. Three-dimensional shapes can be open or closed, narrow or wide, exposing or concealing. This taxonomy of shapes, notice, is closely tied to the activities in which entities with such shapes can engage. An open entity permits movement through it more or less as it is narrow or wide. The same can be said of structural relations between two or more entities: their shapes may be complementary, one entity may be inside, behind or above another, or they may be touching or distant. Again, these relational spatial properties can permit or prohibit activities. So, discovering the structural properties of an entity can often give clues as to the kinds of activities in which it is likely to engage in the next stage of the mechanism. An example is the hint about the nature of replication provided by the double helix structure of DNA (Watson & Crick, 1953a,b).

Activity-enabling properties of entities are not limited to such spatial and structural properties. Entities may also have different kinds of charges and molecules have valences, both of which affect the kinds of bonding activities in which they can engage. Charges have different strengths, different spheres of influence, and different arrangements within the three-dimensional structure of the entity.

In forward chaining, one may ask, what could these entities with these properties in these set-up conditions be expected to do? For example, molecular biologists often attend to the activity-enabling properties of macromolecules, such as their charges, their three-dimensional structures, and the orders of their components. The bases in the double helix of DNA, once the helix is opened, can be expected to guide complementary bonding with other bases. The slight charges on the polar bases are the properties that enable the DNA bases to engage in hydrogen bonding, if charged molecules, complementary to them, are available. So it was reasonable for Crick to conjecture hydrogen bonding between DNA bases and complementary RNA bases within the protein synthesis mechanism. RNA bases were chosen as the complements because empirical work had shown that the mechanism involved RNA at a subsequent stage. Another possible way of forward chaining makes use of another property of the DNA double helix. Gamow used the structure of DNA and possible combinations of the four bases to conjecture the occurrence of a geometrico-mechanical activity (fitting into Gamow holes) that could order the amino acids in proteins.

These examples illustrate the ways that the activity-enabling properties of entities can be fruitfully used in forward chaining to conjecture the activities of the next stage of the mechanism.

4.3.2. Activity consequences

Activities have such properties as rate, duration, strength and sphere of influence. Forward chaining by reasoning about the next stage of the mechanism, one may ask: what is expected of the entities in the subsequent stage, given the prior occurrence of some activity? In a case where hydrogen bonding has occurred between complementarily charged polar molecules, for example, one expects that their charges have been neutralized and are not available to be activity-enabling properties for the next stage. Also, one expects to find a loosely associated (via hydrogen bonding) complex of molecules in that subsequent stage. Because such bonds are weak and easily broken, an even more subsequent stage may involve dissolving them. The hydrogen bonding of a triplet codon on messenger RNA to the anticodon on its corresponding transfer RNA is a transient phenomenon, maintained long enough for the attached amino acid to join the growing polypeptide chain. The hydrogen bond soon dissolves, as the next amino acid is brought into place.

On the other hand, suppose there is evidence for strong covalent bonding, such as finding the larger amount of energy that is required to break the bond. Then one knows that the molecule that is formed by such bonding is stable, and, in any subsequent stage in which the molecule is broken apart, a given amount of energy will be required. Furthermore, in the covalent bonding activity, the valences of the bonding atoms have been used and are not available for additional bonding. The changes are a consequence of the activity of covalent bonding having occurred.

In contrast to these electro-chemical activities, geometrico-mechanical activities, such as lock and key docking, also have characteristic activity consequences. As the docking activity occurs, various stresses and strains may be transmitted through the new, larger structure, possibly changing other active sites, and permitting new geometrico-mechanical activities in the next stage. This sort of change occurs in what is called 'allosteric' regulation, when one molecule docks at one site, and changes the shape of the molecule. (Electro-chemical charges are often also active in allosteric changes.) Such changes may then permit or stop some other activity. In further work on the mechanisms that control protein synthesis, such allosteric changes were found to play an important role (on Monod's discovery of allosteric regulation as the 'second secret of life', see Judson, 1996, Part III).

Knowledge of these general entity-enabling and entity-changing properties of activities allows one to reason forward about the next stage(s) of a mechanism, given knowledge that an activity occurred in a previous stage.

4.3.3. Activity signatures of entities

Now, consider reasoning backward rather than forward. Backtracking may use properties of the entities at a later stage of the mechanism to conjecture the nature of the preceding stages. One asks how such entities could have been produced or what activities could have given rise to, driven, made or allowed this later stage. Decomposing an end product may show its ingredients and provide hints as to what activities could have assembled those ingredients into the product.

As we have seen, beginning with the synthesized protein, biochemists decomposed it into amino acids and found that amino acids are joined by covalent peptide bonds. Once the formation of peptide bonds was shown to be energetically unfavorable, then an energy source was known to be necessary in a preceding step. A search began for high-energy intermediates. These end products carried activity signatures that aided backtracking.

4.3.4. Entity signatures of activities

The characteristic features of an activity may provide clues as to the entities that engaged in it. Distinct kinds of activities require distinct kinds of entities with distinct kinds of properties to produce them. Suppose experimental evidence indicates that weak hydrogen bonding occurs in the mechanism, then one knows that slightly polar molecules must be present. Or, if one conjectures that an activity carries out some role in a mechanism, then that will demand certain properties of entities that can engage in it. Once Crick conjectured that amino acids were ordered via a hydrogen bonding activity, then this stage in the hypothesized mechanism required two types of complementarily, polarly charged molecules. Hence, not only a template RNA was needed, but also an adaptor. When covalent bonding occurs, then entities with appropriate valences must be present. Furthermore, it is likely that enzymes and an energy source are present to carry out this energetically unfavorable reaction.

If lock and key docking has occurred, then two complementarily shaped entities were available. Locks and keys must have complementary structures, of appropriate

sizes, appropriately oriented to each other, and they must come into contact to allow for such geometrico-mechanical activities.

In sum, the discovery of the mechanism of protein synthesis exemplifies the strategy of forward chaining/backtracking, going from entities to activities and back again.

5. Conclusion

Many (perhaps most) of the important discoveries in the biological sciences have been discoveries of mechanisms. So our conclusions about the discovery of the mechanism of protein synthesis will likely generalize to other cases. Centering historical and philosophical analyses on mechanisms reveals previously neglected aspects of scientific discovery, reasoning and theory change. Attending to mechanisms illuminates patterns in the organization of biological knowledge, which have various implications for constructing, evaluating and revising that knowledge over time.

Here we have discussed two discovery strategies (schema instantiation and forward chaining/backtracking) that immediately suggest themselves once one thinks carefully about mechanisms. No doubt, this is just the beginning.

Acknowledgements

This work was supported by the National Science Foundation under grant number SBR-9817942. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation. Lindley Darden's work was also supported by a General Research Board Award from the Graduate School of the University of Maryland, College Park. The characterization of mechanism is due to our work with Peter Machamer, with whom we've had many fruitful discussions. Discussions with Jeff Ramsey about chemical bonding were helpful. Paul Zamecnik kindly supplied reprints for several of his papers. We thank Payman Farsaii and Scott James for research assistance and the following for comments: Chris Cosans, Nancy Hall, Sandra Herbert, Marcia Kraft, Eric Saidel. Sotiris Kotsanis suggested considering how entities and their properties are changed as the mechanism operates. Brad Rives challenged fuzzy metaphysical categories, at least some of which we have sharpened.

References

Bechtel, W. (1984). Reconceptualizations and interfield connections: the discovery of the link between vitamins and coenzymes. *Philosophy of Science, 51*, 265–292.

Bechtel, W. (1986). Introduction: the nature of scientific integration. In W. Bechtel (Ed.), *Integrating scientific disciplines* (pp. 3–52). Dordrecht: Nijhoff.

Bechtel, W. (1988). *Philosophy of science: an overview for cognitive science*. Hillsdale, NJ: Lawrence Erlbaum.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Belozersky, A. N., & Spirin, A. S. (1958). A correlation between the compositions of deoxyribonucleic and ribonucleic acids. *Nature, 182*, 111–112.

Brandon, R. (1985). Grene on mechanism and reductionism: more than just a side issue. In P. Asquith, & P. Kitcher (Eds.), *PSA 1984* (Vol. 2) (pp. 345–353). East Lansing, MI: Philosophy of Science Association.

Brenner, S., Jacob, F., & Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature, 190*, 576–581.

Burian, R. M. (1996). Underappreciated pathways toward molecular genetics as illustrated by Jean Brachet's cytochemical embryology. In S. Sarkar (Ed.), *The philosophy and history of molecular biology: New perspectives* (pp. 67–85). Dordrecht: Kluwer.

Cech, T. R. (2000). The ribosome is a ribozyme. *Science, 289*, 878.

Chadarevian, S. de (1996). Sequences, conformation, information: biochemists and molecular biologists in the 1950s. *Journal of the History of Biology, 29*, 361–386.

Chadarevian, S. de, & Gaudillière, J.-P. (1996). The tools of the discipline: biochemists and molecular biologists, introduction to special issue. *Journal of the History of Biology, 29*, 327–330.

Craver, C. & Darden, L. (2001). Discovering mechanisms in neurobiology: the case of spatial memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 112–117). Pittsburgh, PA: University of Pittsburgh Press.

Crick, F. H. C. (1955). On degenerate templates and the adaptor hypothesis: A note for the RNA tie club. (Unpublished Ms.).

Crick, F. H. C. (1957). Discussion note. In E. M. Crook (Ed.), *The structure of nucleic acids and their role in protein synthesis: biochemical Society Symposium* (Vol. 14) (February 18, 1956) (pp. 25–6). London: Cambridge University Press.

Crick, F. H. C. (1958). On protein synthesis. *Symposium of the Society of Experimental Biology, 12*, 138–167.

Crick, F. H. C. (1959). The present position of the coding problem. *Structure and Function of Genetic Elements: Brookhaven Symposia in Biology, 12*, 35–39.

Crick, F. H. C. (1988). *What mad pursuit: a personal view of scientific discovery*. New York: Basic Books.

Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature, 192*, 1227–1232.

Crick, F. H. C., & Watson, J. D. (1956). Structure of small viruses. *Nature, 177*, 473–475.

Crick, F. H. C., & Watson, J. D. (1957). Virus structure: general principles. In *Ciba foundation symposium on the nature of viruses* (pp. 5–13). London: J. & A. Churchill.

Darden, L. (1991). *Theory change in science: strategies from Mendelian genetics*. New York: Oxford University Press.

Darden, L. (1995). Exemplars, abstractions and anomalies: representations and theory change in Mendelian and molecular genetics. In J. G. Lennox, & G. Wolters (Eds.), *Concepts, theories and rationality in the biological sciences* (pp. 137–158). Pittsburgh, PA: University of Pittsburgh Press.

Darden, L. (1998). Anomaly-driven theory redesign: computational philosophy of science experiments. In T. W. Bynum, & J. Moor (Eds.), *The digital phoenix: how computers are changing philosophy* (pp. 62–78). Oxford: Blackwell.

Darden, L., & Cain, J. A. (1989). Selection type theories. *Philosophy of Science, 56*, 106–129.

Darden, L., & Cook, M. (1994). Reasoning strategies in molecular biology: abstractions, scans and anomalies. In D. Hull, M. Forbes, & R. M. Burian (Eds.), *PSA 1994* (Vol. 2) (pp. 179–191). East Lansing, MI: Philosophy of Science Association.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 44*, 43–64.

Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. *Nature, 173*, 318.

Gaudillière, J.-P. (1993). Molecular biology in the French tradition? Redefining local traditions and disciplinary patterns. *Journal of the History of Biology, 26*, 473–498.

Gaudillière, J.-P. (1996). Molecular biologists, biochemists and messenger RNA: the birth of a scientific network. *Journal of the History of Biology, 29*, 417–445.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49–71.
Gros, F. (1979). The messenger. In A. Lwoff, & A. Ullmann (Eds.), *Origins of molecular biology: A tribute to Jacques Monod* (pp. 117–124). New York: Academic Press.
Gros, F., Hiatt, H., Gilbert, W., Kurland, C. G., Risebrough, R. W., & Watson, J. D. (1961). Unstable ribonucleic acid revealed by pulse labeling of *E. coli*. *Nature*, *190*, 581–585.
Hoagland, M. B. (1955). An enzymic mechanism for amino acid activation in animal tissues. *Biochimica et Biophysica Acta*, *16*, 288–289.
Hoagland, M. B. (1990). *Toward the habit of truth*. New York: Norton.
Hoagland, M. B. (1996). Biochemistry or molecular biology? The discovery of 'soluble RNA'. *Trends in Biological Sciences Letters (TIBS)*, *21*, 77–80.
Hoagland, M. B., Zamecnik, P., & Stephenson, M. L. (1959). A hypothesis concerning the roles of particulate and soluble ribonucleic acids in protein synthesis. In R. E. Zirkle (Ed.), *A symposium on molecular biology* (pp. 105–114). Chicago: Chicago University Press.
Jacob, F. (1988). *The statue within: an autobiography*. New York: Basic Books.
Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*, 318–356.
Judson, H. F. (1996). *The eighth day of creation: the makers of the revolution in biology* (Expanded Edition). Cold Spring Harbour, NY: Cold Spring Harbor Laboratory Press.
Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford, CA: Stanford University Press.
Keyes, M. (1999a). The prion challenge to the 'central dogma' of molecular biology, 1965–1991, Part I: Prelude to prions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *30*, 1–19.
Keyes, M. (1999b). The prion challenge to the 'central dogma' of molecular biology, 1965–1991, Part II: The problem with prions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *30*, 181–218.
Kitcher, P. (1993). *The advancement of science: science without legend, objectivity without illusions*. New York: Oxford University Press.
Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Laudan, L. (1977). *Progress and its problems*. Berkeley: University of California Press.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.
Meheus, J., & Nickles, T. (Eds.). (1999). Scientific discovery and creativity: case studies and computational approaches. Special Issue of *Foundations of Science*, *4*(4).
Morange, M. (1998). *A history of molecular biology* (from the French by M. Cobb, Trans.). Cambridge MA: Harvard University Press.
Nickles, T. (Ed.). (1980). *Scientific discovery, logic and rationality*. Dordrecht: Reidel.
Nickles, T. (Ed.). (1980). *Scientific discovery: case studies*. Dordrecht: Reidel.
Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences (USA)*, *47*, 1588–1602.
Olby, R. (1970). Francis Crick, DNA and the central dogma. In G. Holton (Ed.), *The twentieth century sciences* (pp. 227–280). New York: W. W. Norton.
Pardee, A. B., Jacob, F., & Monod, J. (1959). The genetic control and cytoplasmic expression of 'inducibility' in the synthesis of beta-galatosidase. *Journal of Molecular Biology*, *1*, 165–178.
Pauling, L., & Corey, R. B. (1950). Two hydrogen-bonded spiral configurations of the polypeptide chain. *Journal of the American Chemical Society*, *72*, 5349.
Popper, K. R. (1965). *The logic of scientific discovery*. New York: Harper Torchbooks.
Rheinberger, H.-J. (1997). *Towards a history of epistemic things: synthesizing proteins in the test tube*. Stanford: Stanford University Press.
Rich, A., & Watson, J. D. (1954a). Physical studies on ribonucleic acid. *Nature*, *173*, 995–996.
Rich, A., & Watson, J. D. (1954b). Some relations Between DNA and RNA. *Proceedings of the National Academy of Sciences (USA)*, *40*, 759–764.
Schaffner, K. (1974). Logic of discovery and justification in regulatory genetics. *Studies in History and Philosophy of Science*, *4*, 349–385.

Thieffry, D., & Burian, R. M. (1996). Jean Brachet's alternative scheme for protein synthesis. *Trends in the Biochemical Sciences*, *26*(3), 114–117.

Watson, J. D. (1962). The involvement of RNA in the synthesis of proteins. In *Nobel Lectures in Molecular Biology 1933–1975* (pp. 179–203). New York: Elsevier. Reprinted in 1977.

Watson, J. D. (1965). *Molecular biology of the gene*. New York: W. A. Benjamin.

Watson, J. D. (1968). *The double helix*. New York: New American Library.

Watson, J. D. (2000). *A passion for DNA: genes, genomes and society*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Watson, J. D., & Crick, F. H. C. (1953a). A structure for deoxyribose nucleic acid. *Nature*, *171*, 737–738.

Watson, J. D., & Crick, F. H. C. (1953b). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, *171*, 964–967.

Wimsatt, W. (1972). Complexity and organization. In K. F. Schaffner, & R. S. Cohen (Eds.), *PSA 1972, Proceedings of the Philosophy of Science Association* (pp. 67–86). Dordrecht: Reidel.

Zamecnik, P. C. (1958). The microsome. *Scientific American*, *198*(March), 118–124.

Zamecnik, P. C. (1962a). History and speculation on protein synthesis. *Proceedings of the Symposia on Mathematical Problems in the Biological Sciences*, *14*, 47–53.

Zamecnik, P. C. (1962b). Unsettled questions in the field of protein synthesis. *Biochemical Journal*, *85*, 257–264.

Zamecnik, P. C. (1969). An historical account of protein synthesis, with current overtones — a personalized view. *Cold Spring Harbor Symposia on Quantitative Biology*, *34*, 1–16.

Zamecnik, P. C. (1976). Protein synthesis — early waves and recent ripples. In A. Kornberg, B. L. Horecker, L. Cornudella, & J. Oro (Eds.), *Reflections in biochemistry* (pp. 303–308). New York: Pergamon Press.

Zamecnik, P. C. (1979). Historical aspects of protein synthesis. *Annals of the New York Academy of Sciences*, *325*, 269–301.

Zamecnik, P. C. (1984). The machinery of protein synthesis. *Trends in Biological Sciences Letters (TIBS)*, *9*, 464–466.