What do molecular biologists mean when they say 'structure determines function'?

Gregor P. Greslehner*

University of Salzburg & ERC IDEM, ImmunoConcept, CNRS/University of Bordeaux

October 2018

Abstract

'Structure' and 'function' are both ambiguous terms. Discriminating different meanings of these terms sheds light on research and explanatory practice in molecular biology, as well as clarifying central theoretical concepts in the life sciences like the sequence–structure–function relationship and its corresponding scientific "dogmas".

The overall project is to answer three questions, primarily with respect to proteins: (1) What is structure? (2) What is function? (3) What is the relation between structure and function?

The results of addressing these questions lead to an answer to the title question, what the statement 'structure determines function' means.

*Email: gregor.greslehner@gmail.com

Keywords: philosophy of biology, molecular biology, protein structure, biological function, scientific practice

1 Introduction

'Structure' and 'function' are abundantly used terms in biological findings. Frequently, the conjunct phrase 'structure *and* function' or the directional phrase '*from* structure *to* function' is to be found, indicating that there is a special relation connecting these two concepts. The strongest form of this relation is found in the frequent statement that 'structure *determines* function'. One could easily list several hundreds of references containing such phrases. However, in order not to blow up the references section, I will refrain from doing so. Suffice it to say that biologists make highly prominent use of these concepts in describing their research—molecular biologists, in particular. In this paper, I attempt to clarify these concepts, address their relation, and discuss the role they play in molecular biology's explanatory practice. While these issues can be addressed for many different biological entities on different levels of organization, I restrict the discussion primarily to proteins.

What do biologists refer to when they use this phrase? Is there a particular scientific program or strategy behind the slogan 'structure determines function'? Despite the frequent use of this phrase and the concepts to which it refers, a rigorous analysis is missing. Thus, a philosophical clarification would be a valuable contribution to the conceptual foundations of biology. One such fundamental concept is the sequence–structure–function relationship. "The relationships between sequence, structure, biochemical function and biological role are extremely ill-defined and scant

high quality data are available to allow us to analyse them." (Sadowski and Jones, 2009, 360)

In this paper, I attempt to close this gap by developing an explication of both concepts of *structure* and *function* as they are used in biological practice and discussing which relation holds between them. The third component in this "trinity of molecular biology"—sequence—is the least in need of explication. The standard textbook view holds that sequence determines structure, and structure determines function. I will focus on the second relation.

Without reviewing the rich history of these concepts throughout biology at this point, it is worth noting that functionality and form or structure were thought to be intimately linked from early on. In the early days of biology at the macroscale, the structures had to be observed with the naked eye. Thus, the first examples about the form of bodies or their parts and their functions can be found in physiology and anatomy, for example Harvey's notion of the heart's function to pump blood. From the scale of physiology to the molecular scale, structure and function are closely related. What exactly links these two concepts? Is it a determination relation? And if so, which one is determining the other?

With the invention of microscopes and later the emergence of molecular biology, the structures and functions under consideration shifted from macroscopic entities to individual molecules. In fact, molecular biology put the three-dimensional shape of molecules center stage for explaining biological phenomena. This is the focus of this paper. In particular, the discussion will be confined to the structure and function of *proteins*—with special emphasis on the question whether the former determines the latter.

2 The ambiguity of 'structure'

In a first approximation, 'structure' and 'function' could be interpreted as the most general or neutral way of describing what molecular biologists are doing in their research and what their findings are about. These include mainly the three-dimensional shapes of molecules (or larger cellular structures) and the activities (functions) theses molecules perform in living cells, biochemical pathways, chemical reactions, or just individual steps in such mechanisms. The ultimate aim is to explain biological phenomena with molecular mechanisms, whose entities can be described in physical and chemical terms. The structure of molecules can be described in terms of physics and chemistry—function, however, is a concept that does not appear in physics or chemistry. Let's start by taking a closer look at the notion of structure.

'Structure' is an ambiguous term. Applied to proteins, there is the usual nomenclature of *primary structure* (i.e., a protein's amino acid sequence), *secondary structure* (i.e., common structural motifs like α-helices and β-sheets), *tertiary structure* (i.e., the three-dimensional shape of a single folded amino acid chain), and *quaternary structure* (i.e., the final assembly of a protein if it consists of more than one amino acid chain). Other structurally important components are post-translational modifications and prosthetic groups which are not part of its amino acid composition. All these notions of structure have in common that they are about the molecular composition and shape of a molecule. One meaning of 'structure' denotes the sequence of a polymer, the other meaning is about the three-dimensional shape of a molecule. As will be discussed below, another important ambiguity of 'structure' allows to denote the organization of an interaction network. That leaves us with three different meanings of 'structure':

(1) the sequence of a polymer, (2) the three-dimensional shape of a molecule, and (3) the network organization of several biological entities.

While meanings (2) and (3) are candidates for being functional entities, structure as sequence (1) rather relates the sequences of different polymers (DNA, RNA, and proteins) and also plays a central role in determining the three-dimensional shape of a molecule, structure (2). The primary structure of a protein is just the sequence of amino acids that are put together to form a polypeptide. This amino acid sequence is determined by the corresponding protein-coding gene, which is first transcribed into mRNA and then translated into protein by the ribosome. This scheme is known as the "central dogma of molecular biology":

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{protein}$$

The arrows might be interpreted as determination relations. The textbook view of protein structure and function proceeds as follows:

$$\text{nucleotide sequence} \rightarrow \text{amino acid sequence} \rightarrow \text{protein structure} \rightarrow \text{protein function}$$

Strong evidence supporting the claim that the three-dimensional shape of a protein is determined by the sequence of amino acids alone was provided by the experiments of Christian Anfinsen, showing that ribonuclease could, after treatment with denaturing conditions, regain its form and function (Anfinsen et al., 1961). Later, Merrifield showed that an *in vitro* synthesized sequence of amino acids can carry out the enzymatic activity of ribonuclease, thus gaining its functional form without the aid of any other cellular component (Gutte and Merrifield, 1971). From this and similar experiments, Anfinsen

built general rules of protein folding as a global energy minimum which depends solely on the sequence of amino acids (Anfinsen, 1973). This view is known as "Anfinsen's dogma".

In 1958, John Kendrew's lab determined the first actual three-dimensional form of a protein, myoglobin (Kendrew et al., 1958). The predominant technique to determine protein structures is still X-ray crystallography (Mitchell and Gronenborn, 2017). Other techniques include nuclear magnetic resonance, cryogenic electron microscopy, and atomic force microscopy. X-ray structures in particular have been supporting the view that there is a unique rigid shape—the protein's native, functional state—which would be necessary and sufficient for a protein to carry out its biological function.

To make a long story short, the relation between nucleotide sequence and amino acid sequence has been generally confirmed (although there are much more complicated mechanisms to it, e.g., splicing). However, the part concerning the protein shape and function proves to be much more problematic. That poses a challenge to what Michel Morange calls "the protein side of the central dogma" (Morange, 2006).

To get from amino acid sequence to three-dimensional structure is known as the *protein folding problem*. As the term 'problem' suggests, it poses a serious challenge and remains unsolved to this day. Even though knowledge-based techniques to predict protein structures from their sequence have become impressively sophisticated, successful, and reliable, there are good reasons to suspect that the protein problem might remain unsolved in principle—if the aim is to predict protein folding based on chemical and physical principles only.

Every two years the best prediction tools are tested in a contest, the Critical Assessment of protein Structure Prediction (CASP). Based on experimentally determined structures which are only published after the participants of the contest have

submitted their predictions, the predictions are then compared to the experimental structure. A similar contest for predicting the functions of proteins exists (Critical Assessment of Functional Annotation, CAFA), although it is much less developed. But what is function in the first place?

3 The ambiguity of 'function'

'Function' is also an ambiguous term (Millikan, 1989)—even more so than 'structure'. There is a rich history of debates surrounding different notions of function. The term 'function' has a long tradition in biology and its philosophy (Allen, 2009). Starting with Aristotle, activities in biology were interpreted to *have a purpose*, to be goal-directed (teleological). The standard example is that the heart's function is to pump blood. That the heart also produces noise is not considered to be functional. Classic accounts of function have been predominantly trying to capture the teleological aspect, for example (Wright, 1973). However, intentionality is a problematic notion in biology. In another important account, Robert Cummins (1975) stressed the importance of a component's contribution to the system in which it is contained, rather than why natural selection has favored a certain trait. Although it makes sense in evolutionary biology to have an account of function that captures the evolutionary developments, molecular biology and protein science operate with a different notion of function, i.e., mainly biochemical activity. There seem to be two entirely different questions: What is a structure doing? And how did this structure evolve to do what it does?

Arno Wouters distinguishes four notions of biological function (Wouters, 2003): (1) (mere) activity, (2) biological role, (3) biological advantage, and (4) selected effect.

The last two are issues of evolutionary biology, whereas the former two fall within the molecular biologist's domain. If function is to be determined by a molecule's three-dimensional shape or organization network, only (1) and (2) seem to be the proper reading of 'function' in this context.

Which entities have functions within living organisms? Depending on the level of organization at which one is operating, one could give a different answer: molecules, organelles, cells, tissues, organisms, individuals, populations, ecosystems. The most prevalent candidates in molecular biology are certainly DNA and proteins, although lipids and other biomolecules play important roles in life processes, too.

Traditionally, functions have been attributed to entire genes ("one gene—one enzyme hypothesis"). These views are related to the genetic determinsim view of having a gene for every trait, in which every gene has a function. However, the primary functional units inside a cell are arguably its proteins. Their biochemical activities and biological roles depend crucially on their three-dimensional shapes and network organization, respectively.

One has also to take into account more abstract functional entities, i.e., network modules. These are also called 'structures' but do not refer to the shape of molecules. Its functions ought to be considered as Wouter's second notion (biological role), rather than biochemical activity. "Current 'systems' thinking attributes primary functional significance to the collective properties of molecular networks rather than to the individual properties of component molecules" (Shapiro, 2011, 129). "[A] discrete biological function can only rarely be attributed to an individual molecule [...]. In contrast, most biological functions arise from interactions among many components." (Hartwell et al., 1999, C47). Thus, we can attribute functions as biochemical activities to

individual molecules, wheres systems functions (biological roles) are attributed to organizational structures:

> "Finding a sequence motif (e.g., a kinase domain) in a new protein sheds light on its biochemical function; similarly, finding a network motif in a new network may help explain what systems-level function the network performs, and how it performs it." (Alon, 2003, 1867)

4 Does structure determine function?

Having distinguished between three notions of 'structure' and two notions of 'function', what about the statement 'structure determines function'? Is—in any of its different readings—a certain structure necessary or sufficient for a certain function?

The common textbook view according to Anfinsen has a clear answer: "the central dogma of structural biology is that a folded protein structure is necessary for biological function" (Wright and Dyson, 1999, 322). On first glance, it might appear plausible to assume that a particular structure (understood as molecular shape) is a necessary condition for the proper function of a biological structure (i.e., its biochemical activity). Loss of function is often associated with a loss of the three-dimensional shape of individual proteins. On the other hand, to go for the "sufficient" direction, changes in structures often lead to a decrease in functionality, up to a complete loss. Many diseases for which there are known molecular causes give support to this view. Often it is alterations in the sequence of DNA that result in changed protein shapes that lead to a functionality defect of the organism, which is the definition of a "molecular disease". Alterations of a protein's three-dimensional shape, however, do not necessarily lead to

loss of function. In many cases, changes are "silent", i.e., they don't cause any alteration in phenotype. In rare events, changes might even turn out to be "improvements", which is the driving force of evolutionary development.

However, evidence has been found in the recent years that a significant portion of proteins are intrinsically unstructured in order to be functional, see for example (Forman-Kay and Mittag, 2013). Does the discovery of intrinsically unstructured proteins challenge the relation between structure and function? "[D]isorder aficionados are calling for a complete reassessment of the structure-function paradigm" (Chouard, 2011, 151). Some protein domains fold only upon binding to a suitable target. Others, however, seem to never have an ordered state at all—they remain unstructured even in their functional state.

That a high similarity in sequence does not guarantee a similarity in structure or function has been shown by the Paracelsus Challenge: "a one-time prize of $1000, to be awarded to the first individual or group that successfully transforms one globular protein's conformation into another by changing no more than half the sequence" (Rose and Creamer, 1994, 3). One recent answer to this challenge resulted in the synthesis of two proteins which have 88% sequence identity but a different structure and a different function (Alexander et al., 2007).

Contrary to the view described above, the generalization that a stable three-dimensional structure is necessary or sufficient for a particular function does not hold. It remains true, however, that there is an intimate correlation between structure and function. Prediction tools based on this view are a powerful tool. An attempt to systematically predict the structure and function of proteins based on their amino acid sequence can be found, for example, in (Roy et al., 2010).

To complicate the picture, codon usage is also important: Zhou et al. (2013) have shown that the FRQ protein, which is involved in the circadian clock, is using non-optimal codons, thus translation speed is not optimal. After experimentally optimizing codon usage, the resulting protein—which has the exact same amino acid sequence—folds differently and is no longer functional. This shows that amino acid sequence by itself is not sufficient to determine the three-dimensional structure, let alone its function. In addition to the correct sequence, the folding process has to take place in a certain way which is influenced by the usage of codons and thus the availability of tRNAs, which influences the speed at which the ribosome can proceed translation. Usage of non-optimal codons gives the nascent polypeptide chain some time for the segments that have already been translated to fold in a certain conformation. If translation is too fast, certain intermediate folds which are necessary to reach the final functional conformation can be lost.

Another idea to keep in mind is that evolution operates pragmatically: structures are not the target of selection, functions are. Structures are being re-used for novel functions—there are many biological examples.

If structure does not *determine* function, if a particular structure (in any of its three meanings) is neither necessary nor sufficient for a particular function (in any of its two meanings), may there be another way in which structure and function are related? Perhaps there is a less stringent relationship? I will argue for a supervenience relation (McLaughlin and Bennett, 2018). But before developing this account, we need to clarify which notions of 'structure' and 'function' to use to capture actual scientific practice in molecular biology.

In order to speak about biological functions, a reglemented vocabulary is needed.

The most successful of these is gene ontology (GO) (Ashburner et al., 2000). Fascinating correlation analysis between three-dimensional protein structures from the Protein Data Bank (PDB) and GO terms can be found, for example, in (Hvidsten et al., 2009) and (Pal and Eisenberg, 2005).

According to the textbook picture, there is a linear chain of determination, leading from nucleotide sequences in the DNA via transcription to the nucleotide sequence of RNA, which leads via translation to the amino acid sequence of proteins. The sequence of amino acids, in turn, determines the three-dimensional structure of the protein, whose function, again, is determined by its structure. Given transitivity of this determination relation, one would only need to know the genomic sequence in order to have a complete picture ("blue print") of the functional organism. That is the "holy grail of molecular biology". And like the quest for the holy grail, it is doomed to fail. A strict determination relation does not even hold between the individual pairs.

The reason why the simplified scheme above is still part of the current research "paradigm" lies, on the one hand, in its scientific success: genomics and proteomics have provided unimaginable insights. On the other hand, it fits the mechanistic, reductionistic narrative that has been fashionable in molecular biology. Today, systems biology claims to provide a "holistic" alternative (Green, 2017).

But even without such a strict determination relation between structure and function, both concepts are central to explaining molecular mechanisms in research practice.

In order to understand why molecular biologists explain mechanisms with reference to structure and function, we need to understand what these concepts denote. In a first approximation, molecular biologists analyze a phenomenon by identifying its components that are responsible for the phenomenon in question. These components are the

structures that perform certain biochemical activities, which collectively bring about the phenomenon (biological role). The way in which these entities and their activities are organized is a different meaning of 'structure' which is as important in a mechanistic explanation as individual molecular structures are.

> "Despite the lack of an overarching theory, a Newtonian or quantum mechanics of its very own, molecular biology has become a unifying discipline in virtue of the powers of its techniques, its ability to extrapolate from the molecular to higher levels, and its synthesis of problems of form and function at the molecular level. This synthesis of form and function is a central, ill-understood, and historically important feature of molecular biology." (Burian, 1996, 68)

The ambiguity of the terms 'structure' and 'function' might be useful, for it can be applied to a broad variety of biological research strategies and activities. But, on the other hand, using the term same for different things causes confusion, and the use of metaphorical language might be obscuring certain features and difficulties with this approach.

More recent and thriving approaches in the life sciences have moved beyond the idea that there is a determination relation between structure and function and that by knowing the structure of a protein one could predict its biological function. Today's research in molecular biology is more centered around the *organizational structure* of biological mechanisms. In this way, the ambiguity of the term 'structure' suits to uphold the research slogan, since it can also be applied in a broader sense here than just molecular shapes. The organization of biological systems is the domain of the relatively

new discipline systems biology.

The three-dimensional shape is often a detail that does not contribute to the understanding of a mechanism, but to the contrary would only confuse the mechanistic picture which requires a certain level of abstraction in order to be comprehensive.

But still, how exactly do we get from molecular structures and their (structured) activities to biochemical activities and biological functions? That there might not exist a straightforward mapping from molecular shapes to their biochemical and biological function had been anticipated in the early days of molecular biology:

> "It [molecular biology] is concerned particularly with the *forms* of biological molecules, and with the evolution, exploitation and ramification of these forms in the ascent to higher and higher levels of organization. Molecular biology is predominantly three-dimensional and structural—which does not mean, however, that it is merely a refinement of morphology. It must of necessity enquire at the same time into genesis and function." (Astbury, 1952, 3, original emphasis)

Taking up Francis Crick's remark that "folding is simply a function of the order of the amino acids" (Crick 1958, 144), Morange comments that it is "obviously not a *simple* function" (Morange, 2006, 522). And he observes a semantic change in the meaning of 'function':

> "For Francis Crick, function meant the application of simple rules and principles. For specialists today, function is the result of a complex evolution [...] This shift in the meaning of a word is more than anecdotal. It reflects an active ongoing transformation of biology [...] The mechanistic models of

> molecular biology are no longer considered sufficient to explain the structures and functions of organisms. They have to be complemented and allied with evolutionary explanations" (Morange, 2006, 522).

In order to explain biological phenomena, there is no determination relation that would allow us to track everything down to the chemical and physical properties of proteins, let alone the nucleotide sequences of DNA. Of course, all these issues are relevant to the topic of reduction:

> "if [...] regulatory networks turn out to be crucial to explaining development (and evolution [...]), the reductionist interpretation *may* be in trouble. If network-based explanations are ubiquitous, it is quite likely that what will often bear the explanatory weight in such explanations is the topology of the network rather than the specific entities of which it is composed. [...] How topological an explanation is becomes a matter of degree: the more an explanation depends on individual properties of a vertex, the closer an explanation comes to traditional reduction. The components matter more than the structure. Conversely, the more an explanation is independent of individual properties of a vertex, the less reductionist it becomes." (Sarkar, 2008, 68, original emphasis)

5 Conclusion

Both terms, 'structure' and 'function', are highly ambiguous. So is the widely used conjunct phrase of 'structure and function' that is ubiquitous in biology, as well as the

even stonger claim 'structure determines function'. Perhaps this is why it can be used in many different contexts and for many different explanatory aims in biology. Although providing a certain framework of generality, I argue that a clarification of these concepts is beneficial—for conceptual and philosophical considerations, as well as for the way biologists think about the grand schemes like the "central dogma". Ideally, such an account would also have practical implications and benefit current biological research.

To sum up the results of my analysis, in molecular biology's explanatory practice, 'structure' may refer to:

1. the sequence of polymers,
2. the three-dimensional shape of molecules (or their parts), and
3. the way biological entities are organized.

Of course, different aspects of this distinction play different roles in the explanatory practice with respect to molecular mechanisms. The detailed shape of the interacting molecules is neither necessary nor sufficient for understanding its activities (although correlations are valuable prediction tools before doing experiments in the lab).

The ambiguity of the term 'function' depends on whether the explanation aims at answering the question how a mechanism works or how it came to work that way. Even in the first case one has to distinguish between:

1. the biochemical activity of individual components, and
2. the biological role of network structures.

Whereas biochemical activities of proteins can often be successfully predicted by homology modeling from known molecular shapes, the biological role is rarely an

intrinsic property of an isolated molecule. Rather, the biological role is the mechanistic result of an interaction network of several dynamically interacting molecules.

By comparing the combinatorial possibilities of the different meanings of 'structure' and 'function', a determination relation does not hold between any of them. Instead, I propose a supervenience relation: between the three-dimensional shapes of protein domains and their biochemical activities, and between interaction networks and their biological role. According to my analysis, this is what molecular biologist mean when they say 'structure determines function'.

References

Alexander, P. A., Y. He, Y. Chen, J. Orban, and P. N. Bryan (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proceedings of the National Academy of Sciences 104*(29), 11963–11968. doi:10.1073/pnas.0700922104.

Allen, C. (2009). Teleological notions in biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2009 ed.). `http://plato.stanford.edu/archives/win2009/entries/teleology-biology/`.

Alon, U. (2003). Biological networks: The tinkerer as an engineer. *Science 301*(5641), 1866–1867. doi:10.1126/science.1089072.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science 181*(4096), 223–230. doi:10.1126/science.181.4096.223.

Anfinsen, C. B., E. Haber, M. Sela, and F. H. White, Jr (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences 47*(9), 1309–1314.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock (2000). Gene ontology: tool for the unification of biology. *Nature Genetics 25*, 25–29. doi:10.1038/75556.

Astbury, W. T. (1952). Adventures in molecular biology. In *The Harvey Lectures. Delivered under the auspices of the Harvey Society of New York. 1950–51*, pp. 3–44. Charles C Thomas.

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. Kluwer Academic Publishers.

Chouard, T. (2011). Breaking the protein rules. *Nature 471*, 151–153. doi:10.1038/471151a.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy 72*(20), 741–765. doi:10.2307/2024640.

Forman-Kay, J. D. and T. Mittag (2013). From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure 21*(9), 1492–1499. doi:10.1016/j.str.2013.08.001.

Green, S. (2017). Philosophy of systems and synthetic biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). `https://plato.stanford.edu/archives/sum2017/entries/systems-synthetic-biology/`.

Gutte, B. and R. B. Merrifield (1971). The synthesis of ribonuclease A. *Journal of Biological Chemistry 246*, 1922–1941.

Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray (1999). From molecular to modular cell biology. *Nature 402*(6761 Suppl.), C47–C52. doi:10.1038/35011540.

Hvidsten, T. R., A. Lægreid, A. Kryshtafovych, G. Andersson, K. Fidelis, and J. Komorowski (2009). A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity. *PLoS ONE 4*(7), e6266. doi:10.1371/journal.pone.0006266.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature 181*(4610), 662–666. doi:10.1038/181662a0.

McLaughlin, B. and K. Bennett (2018). Supervenience. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). `https://plato.stanford.edu/archives/spr2018/entries/supervenience/`.

Millikan, R. G. (1989). An ambiguity in the notion "function". *Biology and Philosophy 4*, 172–176. doi:10.1007/BF00127747.

Mitchell, S. D. and A. M. Gronenborn (2017). After fifty years, why are protein X-ray

crystallographers still in business? *The British Journal for the Philosophy of Science 68*(31), 703–723. doi:10.1093/bjps/axv051.

Morange, M. (2006). The protein side of the central dogma: Permanence and change. *History and Philosophy of the Life Sciences 28*(4), 513–524.

Pal, D. and D. Eisenberg (2005). Inference of protein function from protein structure. *Structure 13*, 121–130. doi:10.1016/j.str.2004.10.015.

Rose, G. D. and T. P. Creamer (1994). Protein folding: Predicting predicting. *PROTEINS: Structure, Function, and Genetics 19*, 1–3.

Roy, A., A. Kucukural, and Y. Zhang (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols 5*(4), 725–738. doi:10.1038/nprot.2010.5.

Sadowski, M. and D. T. Jones (2009). The sequence–structure relationship and protein function prediction. *Current Opinion in Structural Biology 19*, 357–362. doi:10.1016/j.sbi.2009.03.008.

Sarkar, S. (2008). Genomics, proteomics, and beyond. In S. Sarkar and A. Plutynski (Eds.), *A Companion to the Philosophy of Biology*, pp. 58–73. Blackwell Publishing Ltd.

Shapiro, J. A. (2011). *Evolution: a view from the 21st century*. FT Press Science.

Wouters, A. G. (2003). Four notions of biological function. *Studies in History and Philosophy of Biological and Biomedical Sciences 34*(4), 633–668. doi:10.1016/j.shpsc.2003.09.006.

Wright, L. (1973). Functions. *The Philosophical Review 82*(2), 139–168. doi:10.2307/2183766.

Wright, P. E. and H. J. Dyson (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology 293*, 321–331. doi:10.1006/jmbi.1999.3110.

Zhou, M., J. Guo, J. Cha, M. Chae, S. Chen, J. M. Barral, M. Sachs, and Y. Liu (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature 495*, 111–115. doi:10.1038/nature11833.