MIND-BODY PROBLEM

See **Consciousness; Intentionality; Physicalism; Supervenience**

MOLECULAR BIOLOGY

The term 'molecular biology' was introduced by Warren Weaver in 1938 in an internal report of the Rockefeller Foundation: "And gradually there is coming into being a new branch of science—molecular biology . . . in which delicate modern techniques are being used to investigate ever more minute details of certain life processes" (as quoted in Olby 1974, 442). What Weaver may have only dimly foreseen is that these new techniques would ultimately transform the practice of biology in a way comparable only to the emergence of the theory of evolution in the nineteenth century. At the beginning of the twenty-first century, molecular biology has become most of biology, either *constitutively,* insofar as biological structures are characterized at the molecular level as a prelude for further study, or at least *methodologically,* as molecular techniques have become a preferred mode of experimental investigation of a domain. Recent biological work at the organismic and lower levels of organization (cytology, development, neurobiology, physiology, etc.) increasingly fall under the former rubric. Work in demography, epidemiology, and ecology falls under the latter, with ecology perhaps being the subdiscipline within biology that has most resisted molecularization. Work in evolution falls under both: constitutively, when the evolution of molecules and molecular structures forming organisms is studied for its own sake, and methodologically, when molecular techniques (most notably, DNA sequencing) are used to reconstruct evolutionary history. This article will be largely restricted to the constitutive aspect of molecular biology, since that is what has so far (perhaps deservedly) commanded most philosophical attention.

The decade following Weaver's introduction of molecular biology saw the steady increase in the use of "delicate" molecular techniques, in particular, x-ray crystallography, to study biological macromolecules "minutely," increasingly with an emphasis on proteins. The central problem was the elucidation of the three-dimensional structures (the relative positions of the atoms) of biological macromolecules. The structure of proteins was supposed to explain their behavior. Proteins were singled out because they were believed to be the most important of these macromolecules. In particular, since the establishment of biochemistry as a discipline in the 1920s, enzymes and their interactions had been held to be the key to understanding *metabolism* (the catchall term for the complex chemical reaction systems that characterize life). All enzymes are proteins. Until the early 1940s, it was believed that the hereditary material (the genes) was also likely to be proteins. The nucleic acids, constructed out of only four nucleotide base types (adenosine [A], cytosine [C], guanine [G], and thymine [T]), were believed to be insufficiently complex to be able to specify the immense variety of known genes.

However, experimental work starting in the early 1940s showed that the hereditary substance—specifying 'genes' (see Genetics)—was deoxyribonucleic acid (DNA). Attention then shifted to deciphering the physical structure of DNA, a problem that was solved by Watson and Crick (1953)

with their double-helix model. The construction of this model and its subsequent confirmation marks a development of signal importance for modern biology (Sarkar 2005, Ch. 1). It ushered in the "classical" age of molecular biology (see the next two sections) with an intriguing informational interpretation of biology (see Biological Information). Important conceptual innovation also came from Monod and Jacob in the early 1960s, who constructed the allostery model to explain cooperative behavior in proteins and the operon model of gene regulation (Monod 1971; Jacob 1973; see below). Genes were interpreted as DNA sequences either specifying proteins (the *structural* genes) or controlling the action of other genes (the *regulatory* genes). Perhaps the most important development in classical molecular biology was the establishment of a genetic "code" delineating the relation of DNA sequences to amino acid residue sequences in proteins. (Both DNA and protein are linear molecules in the sense that they consist of units connected in a chain through strong [covalent] chemical bonds.) Gene *expression* took place by the *transcription* of DNA to ribonucleic acid, RNA, at the chromosomes (in the nucleus), and the *translation* of these transcripts into protein at the ribosomes (in the cytoplasm). The one gene–one enzyme credo of classical genetics was transformed into the one DNA segment–one protein chain credo of molecular biology (see Genetics).

Crucial to the program of molecularizing biology was the expectation—first explicitly stated by Waddington (1962)—that gene regulation explained tissue differentiation and, ultimately, morphogenesis in complex organisms. Genetic reductionism, the thesis that genes alone can explain organismic features, long predates molecular biology (Sarkar 1998). However, the molecular interpretation of the gene allowed the general explanatory success of molecular biology to be co-opted as a success of molecular genetics. In such a context, Waddington's thesis was positively received and helped usher in an era dominated by *developmental genetics*, according to which organismic development was to be understood through the action of genes. Mayr (1961) and others introduced the metaphor of the genetic program to characterize the putative relation between genomic DNA and organismic features. As molecular genetics began to dominate the research agenda of molecular biology in the 1970s, the emergence of organismic features came to be viewed as determined by "master control genes" (Gehring 1998). This view was initially supported by the demonstration that some DNA sequences (such as the homeobox) were conserved across a wide variety of species. DNA came to be viewed as the molecule "defining" life, a view that helped initiate the massive genome sequencing projects of the 1990s, which were supposed to produce a gene-based complete biology that delivered on all the promises of molecular developmental genetics. In general, because of the presumed primacy of DNA in influencing organismic features, starting in the early 1960s, molecular genetics began to dominate research in molecular biology.

Genetics and development were the earliest biological subdisciplines to be redefined by molecular biology. In the case of evolutionary biology, as early as the 1950s, Crick (1958) pointed out that the genotype/phenotype relation could be reinterpreted as the relation between DNA and protein, with proteins constituting the subtlest form of the expression of a phenotype of an organism. Consequently, the evolution of proteins (and, later, DNA sequences), especially the question of what maintained their diversity within a population, became a topic of investigation—in the 1960s, these studies led to the neutralist challenge to the received view of evolution (see Evolution). More importantly, changes at the level of DNA sequences, provided that these were selectively neutral, permitted the construction of a "molecular clock" that could be used to reconstruct evolutionary history more accurately than could be achieved by traditional morphological methods (see Population Genetics).

Meanwhile, biochemistry and immunology were reconstituted by the new molecular biology in ways that were not unexpected. That enzyme interactions and specificity would be explained in molecular terms was no surprise (see "Classical Molecular Biology" below). However, immunological specificity was also believed to be explainable by the same mechanism. This model of immune action was coupled to a selectionist theory of cell proliferation to generate the clonal theory of antibody formation, which combined molecular and cellular mechanisms in a novel fashion (see Immunology). In both biochemistry and immunology, what was largely at stake was the development of models that could explain the observed specificity of interactions: Enzymes reacted with only very few substrates; antibodies were highly specific to their antigens.

By the late 1970s, it became clear that the simplicity of the picture of genetics inherited from the 1960s was being lost. The initial picture was generated from an exploration of the genomes of prokaryotes (single-celled organisms without a nucleus), especially the bacterium *Escherichia coli*. In prokaryotes, every piece of DNA has a structural

or regulatory function. In the 1970s, it was discovered that the genetics of eukaryotes (organisms with cells with nuclei) turned out to have an unexpected complexity. In particular, large parts of the genomic DNA sequences apparently had no function: These segments of "junk" DNA were interspersed between genes on chromosomes and also within genes. After RNA transcription, noncoding segments within genes were *spliced out* before translation. Gene regulation in eukaryotes was qualitatively different and more complicated than in prokaryotes. Some organisms used nonstandard genetic codes and other alternatives (see Sarkar 1996 for a detailed account.)

Subsequent work in molecular biology has only added to this picture of complexity, so much so that it is reasonable to suspect that the classical picture is breaking down. RNA transcripts are subject to *alternative splicing*, with the same DNA gene corresponding to several proteins. RNA is edited, with bases added and removed, before translation at the ribosome, to such an extent that it is sometimes difficult to maintain that some gene actually does code for a given protein. There is no obvious relation between the number of genes in an organism and its morphological or behavioral complexity. Most importantly, it now appears that a fair amount of the DNA thought to be junk is transcribed into RNA though not translated. Thus, presumably, much of the so-called junk DNA is functional, though the nature of these functions remains controversial.

This article will concern both classical molecular biology and the postgenomic molecular biology of the modern era. It will not only discuss issues in the philosophical interpretation of the classical era, which are fairly well characterized, but also include more speculative discussions of issues raised by recent developments.

Classical Molecular Biology

Classical molecular biology can be viewed in continuity with both the genetics and the biochemistry of the era that preceded it. From biochemistry—in particular, the study of enzymes in the 1920s and 1930s—early molecular biology inherited the mechanistic proposal that the function or behavior of biological molecules was "determined" by its structure, an idea that went back to Ehrlich's "side-chain" theory in the late nineteenth century. In the 1950s, structural modeling of biological macromolecules, especially proteins, was pioneered by Pauling and his collaborators using data from x-ray crystallography (see, e.g., Pauling and Corey 1950). By the early 1960s, a handful of such structures were fully solved. These structures, along with the structure of DNA, seemed to confirm the hypothesis that structure explains behavior. Perhaps more surprisingly, it was found that structural interactions seemed to be mediated entirely by the shape of active sites on molecules and that the sensitive details of structure and shape were maintained by very weak interactions.

These experimental observations led to four seemingly innocuous rules about the behavior of biological macromolecules, which in the 1960s and 1970s formed the theoretical core of molecular biology (Sarkar 1998, 149–150):

1. The weak interactions rule: The interactions that are critical in molecular processes are very weak.
2. The structure-function rule: The behavior of biological macromolecules can be explained from their structures, as determined by techniques such as crystallography.
3. The molecular shape rule: These structures, in turn, can be characterized entirely by molecular size, external shape (especially), and some general properties (such as hydrophobicity) of the different regions of the surfaces;
4. The lock-and-key fit rule: In molecular interactions, molecules interact only when there is a lock-and-key fit between the two molecular surfaces. There is no interaction when these fits are destroyed.

A lock-and-key-fit thus based on shape is an obvious way of achieving stereospecific capacity, thus resolving the critical problem for classical molecular biology. Because they are most intimately involved in the explanation of specificity, the molecular shape and lock-and-key-fit rules are the most important in this respect. In what follows, these will be called the rules of classical molecular biology.

In the 1960s and 1970s, these rules were deployed with remarkable success. As noted earlier, enzymatic and immunological interactions were among those that were immediately brought under the aegis of the new molecular biology. Two other cases are even more philosophically interesting:

- The allostery model explains why some molecules such as hemoglobin show *cooperative* behavior. In the case of hemoglobin, there is a nonlinear increase in the binding of oxygen after binding is first initiated. This is explained by conformational—shape—changes in the molecular subunits of hemoglobin; and

- The operon model explains *feedback*-mediated gene regulation in prokaryotes: The presence of a substrate activates the production of a protein that interacts with it, and its absence inhibits that production (see Monod 1971 for an accessible accurate account of these two examples and a conceptual summary of theoretical reasoning in early molecular biology).

Both cooperativity and feedback phenomena formed part of the traditional repertoire of holists in biology (see the next section, which will discuss the philosophical significance of the success of such structural explanation in molecular biology).

However, the 1950s also saw the elaboration of a radically different model of biological specificity, based on the concept of *information*, which was introduced into genetics only in 1953 (Sarkar 1996). This concept soon came to play a foundational role in molecular genetics. DNA was supposed to be the repository of biological information, a genetic "program" was supposed to convert this information into the adult organism, and new information was supposed to result from random mutation (and be maintained by selection) and never incorporated into the genome from the environment. Crick (1958) enshrined these assumptions in what he called the central dogma of molecular biology:

> This states 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. (153) (emphasis in original)

Information, according to Crick, was the sequence of nucleotide bases in DNA or the sequence of amino acid residue in protein molecules. Note the contrast here with the stereospecific physical model of specificity. The dogma has continued to be an important regulative principle of molecular biology in the sense that it is presumed for further theoretical reasoning. Whether it survives recent developments will be discussed later in this essay.

However, the complexities of eukaryotic genetics, as discovered in the 1970s and 1980s, already began to challenge the central dogma (but see Thiéffry and Sarkar 1998). Much of this work was made possible by the development of technologies based on the polymerase chain reaction in the 1980s. There were five salient discoveries that challenged the simple picture inherited from prokaryotic genetics (Sarkar 2005, Ch. 8) (see Genetics):

1. The genetic code is not fully universal, the most extensive variation being found in mitochondrial DNA in eukaryotes. However, there is also some variation across taxa (see Fox 1987 for a review).
2. DNA sequences are not always read sequentially in blocks. There are overlapping genes, genes within genes, and so on (Barrell, Air, and Hutchison 1976). Thus, two or more different proteins could be specified by the same gene.
3. As noted earlier, not all DNA in the genome is functional. Intervening sequences—within and between structural genes—must be spliced out from transcripts (Berget, Moore, and Sharp 1977; Chow et al. 1977). This discovery helped resolve the so-called C-value paradox (Cavalier-Smith 1978), that is, the absence of any obvious correlation between the size of the genome and the morphological and behavioral complexity of an organism.
4. The same transcript may be spliced in different ways (Berk and Sharp 1978). One consequence of such alternative splicing is that, as with overlapping genes, two or more different proteins could be specified by the same gene.
5. Besides splicing, RNA is sometimes subject to extensive editing before translation at the genome (Cattaneo 1991).

These developments have led to skepticism of the relevance of the coding model of the DNA/protein relationship and of the informational model of specificity (see the next section). Though philosophers (and some biologists) have been slow to recognize this, the credo of one DNA segment–one protein chain has long become irrelevant in molecular biology. The modern era presents even more significant challenges, as later sections of this essay will underscore.

Philosophical Interpretations

Philosophy of biology only emerged as a recognizable part of philosophy of science only in the late 1960s. In the early years, considerable attention was paid to molecular biology, especially with respect to the issue of reductionism, but starting in the late 1970s, attention within philosophy of biology began to be concentrated solely on evolutionary theory, much to the detriment of the field. Attention shifted back to molecular biology in the 1990s, with some work now being done on the question of biological information besides reductionism. Since then, classical molecular biology has

been increasingly scrutinized by philosophers, though not as much as it deserves. This section will focus on reduction and information. However, important philosophical work has also been done on other forms of conceptual change in molecular biology and, lately, experimentation in the field (Culp 1995; Rheinberger 1997).

Reduction

The first question about molecular biology that interested philosophers was whether it could be interpreted as a reductionist enterprise in the same way as the kinetic theory of matter was reductionist within classical physics (see Reductionism). The model of reduction then in vogue was due to Nagel (1961) with some modification by Schaffner (1967): It viewed reduction as a deductive-nomological explanation but with the reduced laws as the explananda (see Explanation; Nagel, Ernest). The debate soon centered on the question of whether molecular genetics was reducing or replacing Mendelian genetics. While Schaffner (1967) made the case for successful reduction, this position was attacked by Hull (1972) on the grounds that molecular biology did not have laws and theories (as logical empiricists envisioned those entities). Subsequently, an antireductionist consensus developed (also influenced by Kitcher [1984]).

This consensus was subsequently challenged by Sarkar (1989 and 1998), Waters (1990), and others, but only by rejecting the Nagel-Schaffner formal model as being relevant to substantive questions about reduction. (Even earlier, Wimsatt [1976] had argued against the relevance of the Nagel-Schaffner model.) In these analyses, what is at stake is that properties of wholes are being explained by properties of parts interacting locally. The allostery and operon models are philosophically critical exemplars of this approach because the former explains cooperativity and the latter feedback, both of which formed part of the conceptual repertoire of traditional holists (see Emergence). Enzymatic and immunological specificity provide more mundane examples. However, most of these cases are much simpler than that of providing fully reductionist explanations of quintessentially Mendelian genetic phenomena such as the segregation or assortment of alleles. In these cases—central to the question of reducing Mendelian genetics to molecular genetics—reductionist explanation remains piecemeal and, in many ways, incomplete. However, there is every reason to believe that the relevant lacunae will be filled without requiring new conceptual or theoretical resources.

Nevertheless, even during the classical era, a few anomalies remained, though none serious enough to call into question the viability of the reductionist project. In particular, there has never been a successful parts-whole account of dominance (that is, the dominance of one trait or allele over another) (see Genetics). There is some reason to believe that explaining dominance at the molecular level will require appeal to topological properties of networks, but such a move would take explanation beyond the reductionist realm (see "Philosophical Speculations" below).

Information

Though it is commonplace to talk of biological information, no successful formal definition of the concept in the context of molecular biology has ever been given. Because of difficulties that the concept of information encountered in the late 1980s and 1990s, this failure led Sarkar (1996) to suggest that information in molecular biology was a metaphor masquerading as a theoretical concept (Griffiths 2001) (see Scientific Metaphors). For Crick (1958), information consisted of sequences, of DNA or protein. Informally, this is what 'information' is probably taken to mean in most contexts. The first point to note is that any such definition would require that the concept of information being used *not* be Shannon's (1948) communication-theoretic notion of information, which requires the estimation of the frequency of symbols drawn from a set. Thus, mathematical information theory based on Shannon's concept simply becomes irrelevant in this context (for a contrary position, see Yockey 1992). At the very least, any usable concept of biological information must refer to individual sequences and be symbolic, semantic, or semiotic, in the sense that it must capture the idea that the sequence is a "sign" for something else (Sarkar 2005, Ch. 10). As such, it must account for biological specificity.

The concept was central to two related theoretical interpretations within molecular biology:

(a) that the DNA/protein relation is a genetic code, typically extended to suggest that all phenotypic traits are encoded in the DNA of the genome; and
(b) that the genome constitutes a genetic program for the organism.

As discussed earlier, developments within eukaryotic genetics began to limit the scope of the genetic code in the 1970s. Any claim of the existence of a genetic program at the very least constitutes a claim of genetic reductionism, and at the

very worst a claim of genetic determinism. Genetic reductionism must be clearly distinguished from physical reductionism (the physical explanation of properties of wholes from properties of parts, which was discussed earlier). Genetic reductionism is the claim that organismic features are satisfactorily explained by appeal to properties of genes or DNA (without recourse to properties of other molecules). It is, for instance, central to the project of developmental genetics. Such a reductionism was never very plausible; consequently, the metaphor of a genetic "program" was always troubled (Keller 2000). Nevertheless, the "program" metaphor was quite influential during the heyday of developmental genetics. As the next two sections will underscore, it does not survive even in a mitigated form in the postgenomic era. The failure of genetic reductionism makes any stronger claim of genetic determinism irrelevant.

Viewing information as sequence, Crick (1958) also proposed the *sequence hypothesis:* that the sequence of amino acid residues in a protein (also called its *primary* structure) determines its three-dimensional conformation (also called its *tertiary* structure). Attempting to show how this comes about came to be called the protein folding problem. It has never been successfully solved, and not for lack of effort (Sarkar 1998). Moreover, for many proteins, it is known that sequence alone is insufficient for specifying three-dimensional conformation. It may even be the case that the same sequence can lead to several different conformations. This failure casts additional doubts on the utility of the concept of biological information stored in the genome, at least in the sense Crick intended it. Even if the genetic code were as exceptionless and predictively successful as was believed in the 1960s, all that it would allow is the inference of an amino acid residue sequence from the DNA. If the protein sequence does not determine its conformation, ipso facto, the DNA sequence cannot. It follows that the information within the DNA cannot specify phenotypes even further removed from the genome.

To the extent that the genetic code still remains useful, a proper explication of the concept of biological information remains an unaccomplished philosophical task of some importance (see Biological Information).

The Modern Era

By the "modern era" of molecular biology is meant the period beginning with the production of large genomic sequences in the 1990s. It is also referred to as the Genomic, Postgenomic, or, less accurately, Proteomic era ("proteomic" is less accurate because, to date, there has been limited progress in proteomics; see below). What marks this era is the study of large genomic sequences, and not individual alleles that had been previously identified by their phenotypic effects.

Genomics and Postgenomics

Genomics was ushered in by the decision to sequence the entire human genome as an organized project (the Human Genome Project [HGP]), involving a large number of laboratories in the late 1980s. Subsequently, similar projects were established to sequence the genome of many other species. To date, genomes of over 150 species have been sequenced. Almost every month sees the announcement of the completion of sequencing for a new species. The sheer volume of sequence information that has been produced has spawned a new discipline of "bioinformatics" dedicated to the computerized analyses of biological data.

When the HGP was first proposed, there was considerable controversy among biologists about its wisdom (Tauber and Sarkar 1992; Cook-Deegan 1994). There were:

(i) doubts about its ability to deliver on the bloated promises made by proponents of its scientific and, especially, medical benefits;
(ii) questions whether such organized "Big Biology" projects were wise science policy because of their potential effect on the ethos of biological research; and
(iii) worries that society would be legally and medically ill-prepared to cope with the results of sequencing that came too rapidly, in contrast to the normal slower accumulation of human genomic sequence information. It was feared that legislation protecting genetic privacy and preventing genetic discrimination would not be in place; there would be a shortage of genetic counselors; and so on.

In one important respect, the critics were correct: There have been few immediate medical benefits from the HGP, and no significant such innovation seems forthcoming. Instead, recent work underscores the importance of gene/environment interactions that critics had routinely invoked to criticize the claims of the HGP (see Heredity and Heritability). However, in another sense, even the most acerbic critics should now accept that the scientific results of the sequencing projects, taken together, have been breathtaking.

Contrary to the expectations of the HGP's proponents, few successful predictions about organismic development have come from sequence information alone (Stephens 1998). However, genomic research is persistently throwing up surprises:

1. The most important surprise from the HGP is that there are probably only about 30,000 genes in the human genome, compared with an estimate of 140,000 as late as 1994 (Hahn and Wray 2002). In general, plant genomes are expected to contain many more genes than the human genome. Morphological or behavioral complexity is not correlated with the number of genes that an organism has. This has been called the G-value paradox (ibid).
2. The number of genes is also not correlated with the size of the genome, as measured by the number of base pairs. The fruit fly *Drosophila melanogaster* has 120 million base pairs but only 14,000 genes; the worm *Caenorhabditis elegans* has 97 million base pairs but 19,000 genes; the mustard weed *Arabidopsis thaliana* has only 125 million base pairs and 26,000 genes; while humans have 2,900,000,000 base pairs and 30,000 genes (Hahn and Wray 2002).
3. At least in humans, the distribution of genes on chromosomes is highly uneven. Most of the genes occur in highly clustered sites. Most of these genes are expressed in many tissues—the so-called "housekeeping" genes (Lercher, Urrutia, and Hurst 2002). However, the spatial distribution of cluster sites appears to be random across the chromosomes. (Cluster sites tend to be rich in C and G, whereas gene-poor regions are rich in A and T.) In contrast, the genomes of arguably less complex organisms, including *D. melanogaster*, *C. elegans*, and *A. thaliana*, do not have such pronounced clustering.
4. Only 2% of the human genome codes for proteins, while 50 % of the genome is composed of repeated units. Coding regions are interspersed by large areas of noncoding DNA. However, some functional regions, such as *HOX* gene clusters, do not contain such intervening sequences.
5. Scores of genes appear to have been horizontally transferred from bacteria to humans and other vertebrates, though apparently not to other eukaryotes. However, this issue remains highly controversial.
6. Once attention shifts from the genome to the proteome, or the protein complement of a cell (see below for more detail), a strikingly different pattern emerges. The human proteome is far more complex than the proteomes of the other organisms for which the genomes have so far been sequenced. According to some estimates, about 59% of the human genes undergo alternative splicing, and there are at least 69,000 distinct protein sequences in the human proteome. In contrast, the proteome of *C. elegans* has at most 25,000 protein sequences (Hahn and Wray 2002).
7. It now appears that noncoding DNA is routinely transcribed into RNA but not translated in complex organisms (Mattick 2003). It seems that these RNA transcripts form regulatory networks that are critical to development. Interestingly, the amount of noncoding DNA sequences in organisms appears to grow monotonically with the morphological complexity of organisms.
8. At least in *A. thaliana*, there is evidence of genome-wide non-Mendelian inheritance during which specifications from the grand-parental, rather than parental, generation are transmitted to organisms (Lolle et al. 2005).

An important task of modern molecular biology is to make sense of these disparate unexpected discoveries. One conclusion seems unavoidable: Any concept of the gene reasonably close to that in classical genetics will be irrelevant to the molecular biology of the future (see Genetics).

Proteomics

The term "proteome" was introduced only in 1994 to describe the total protein content of a cell produced from its genome (Williams and Hochstrasser 1997). Unlike the genome, the proteome is not even approximately a fixed feature of a cell (let alone an organism), but changes over time during development. Deciphering the proteome, and following its temporal development during the life cycle of each tissue of an organism, has emerged as the major challenge for molecular biology in the postgenomic era. This project has been encouraged by the discovery of unexpected universality of developmental processes at the level of cells and proteins (Gerhart and Kirschner 1997). For instance, even though hundreds of genes are known to specify molecules involved in transport across cellular membranes, there are only about twenty transport mechanisms in all living systems. The emergence of proteomics in the wake of the various sequencing projects signals an acceptance of the position that studying

processes largely at the DNA level will not suffice to explain phenomena at the cellular and higher levels of organization. Even genomics did not go far enough; a sharper break with the past will be necessary.

Nevertheless, in one very important sense, the emergence of proteomics recaptures the spirit of early molecular biology, when all molecular types, but especially proteins, were the foci of interest, and the deification of DNA had not replaced a pluralist vision of the molecular basis for life. In the late 1960s, Brenner and Crick proposed "Project K, the complete solution of *E. coli*." *E. coli* (strain K-12) was selected as a model organism because of its simplicity (as a unicellular prokaryote) and ease of laboratory manipulation. Project K included: (i) a "detailed test-tube study of the structure and chemical action of biological molecules (especially proteins)"; (ii) completion of the models of protein synthesis; (iii) work on the structure and function of cell membranes; (iv) the study of control mechanisms at every level of organization; and (v) the study of the behavior of natural populations, including population genetics. Once *E. coli* was solved, and biology was supposed to move on to more complex organisms (Crick 1973, 67).

Notice that DNA receives no preferential attention at the expense of other molecular components in Project K and that the centrality of proteins as the most important active molecules in a cell is recognized. Project K accepts that there is much more to the cell than DNA; it accepts that no simple solution of the cell's behavior can be read from the genomic sequence. After a generation of infatuation with DNA and genetic reductionism, the aims of proteomics return in part to the vision of biology incorporated in Project K. However, at least in one important way, that project went even beyond proteomics as currently understood: It emphasized all levels of organization, whereas the explicit aims of proteomics are limited to the protein level. The future will probably require further expansion.

Meanwhile, work on proteins has also generated unexpected challenges. In particular, the four rules of classical molecular biology have not survived intact, and at least the last three will require some modification. It now appears—though the essential idea goes back to the 1960s—that the fit between interacting sites of protein molecules is more dynamic than in the classical model, with the active site often "inducing" an appropriate fit (see e.g., Koshland and Hamadani 2002). It also appears that a more complicated model than the original allostery model will be required to account for many cases of cooperativity. A systematic philosophical appraisal of these developments is yet to be undertaken.

Philosophical Speculations

The developments described in the last section are so recent that any attempt to interpret their philosophical significance must remain partly speculative. Some of the empirical generalizations noted will undoubtedly be challenged by further work in the near future—if the recent past of molecular biology is any guide to its future. Moreover, there has been very little philosophical attention to these developments.

Beyond Reduction?

That the four rules of classical molecular biology are being challenged, at least to some extent, is not reason enough to generate any new skepticism about the reductionist interpretation of explanation in molecular biology. They do not bring the physical explanation of wholes by parts into question. However, if an RNA-based (or other) regulatory network turns out to be crucial to explaining development (and evolution, as Mattick 2003 argues), 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. As noted earlier, some classical phenomena such as dominance have already been known to resist straightforward reductionist explanation (Sarkar 1998).

Topological explanations have not received the kind of attention from philosophers they deserve, even though networks have lately entered the center stage of scientific attention (Mattick and Gagen 2005). Here "topology" refers to the connectivity properties of systems such as networks, which, without loss of generality, can be modeled as directed graphs. The vertices of such a graph represent components of a system, and edges (between vertices), with appropriate directionality and weights, represent interactions between such vertices. 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) (see Reductionism). Conversely, the more an explanation is independent of individual properties of a vertex, the less reductionist it becomes. In the latter case, if explanations invoke properties of a graph that measure its connectivity, then these are

topological explanations. Such connectivity measures include the number of edges in the graph, the distribution of edge degree between vertices (the "degree" of a vertex being the number of edges incident on it), and so on. (For a review of network theory, see Newman 2003). If topological explanations become necessary in molecular biology, it will mark a serious philosophical break with the reductionist classical era.

Beyond DNA Information?

As noted earlier, there is as yet no fully satisfactory account of biological information that is appropriate for molecular biology. However, the developments within eukaryotic genetics and, especially, genomics strongly suggest that the view that DNA is the sole carrier of information, however it is characterized, cannot be sustained at least for organisms more complicated than prokaryotes and perhaps not even for them. Most of the critical interactions that determine the future behavior of a cell seem to occur at the level of RNA: splicing, RNA editing, and so on. Because of this feature of cellular interactions, Sarkar (2005, Ch. 14) has speculated that the DNA genome consists of a relatively static set of sequestered modular templates (resulting in the "SMT model" of the cell), far from the classical view of the genome coding a program for development. The failure of the sequence hypothesis for many proteins only increases skepticism about the classical picture.

The routine generation of untranslated RNA transcripts from the genome also suggests that should cellular processes be viewed informationally, RNA networks form a parallel information-processing system partly independent from the genomic DNA (Mattick 2003). At present, it is unclear whether such information must also be viewed semiotically, though it seems likely, since the simplest way in which RNA sequences can be viewed as carriers of information is by the specification of information by the RNA sequences.

Similarly, the discovery of ubiquitous non-Mendelian genetic specification in *A. thaliana* (Lolle et al. 2005) also suggests that there is yet another parallel system of heredity that can also potentially be viewed informationally and, once again, is not specified through DNA. It is also possible that all such phenomena are best interpreted not informationally but using the more traditional—generally structural—conceptual apparatus of physics and chemistry. However, the distinction between the two frameworks becomes blurred in the case of RNA because the relation between the sequence and three-dimensional conformation seems to be relatively straightforward, at least much more so than in the case of proteins.

Finally, in these discussions of biological information, two issues should be distinguished:

- whether an informational framework for molecular biology is of any use; and
- whether, within any such framework, DNA (or, more restrictively, genomic DNA) is the sole repository of that information.

The problems mentioned here provide a forceful argument against the second claim, leaving open the status of the first.

Toward a Dynamic Account of the Organism

One problem with informational interpretations of molecular biology is that they have always been static: Time does not enter explicitly into accounts of biology based on the transfer of information, though, implicitly, such transfer must take place during some time interval. Recall that the proteome is not a static feature of the organism, let alone the cell: Proteomics requires a commitment to the characterization of cellular and organismic change over time. Moreover, the recent discoveries of potentially ubiquitous RNA network-based regulation also underscore the importance of dynamic accounts explicitly taking time into consideration. Moreover, new microarray techniques and their extensions are increasingly making temporal stages of cellular changes empirically accessible. The challenge remains to develop a theoretical framework to interpret the empirical information.

Any such framework can begin with either a physicalist or an informational characterization of cellular processes or a mixture of both, though prospects for a physicalist account do not seem particularly promising because of the sheer complexity of the molecular networks involved (Sarkar 2005, Ch. 10). But a dynamic informational account also leads to uncharted territory. In retrospect, what seems surprising is how successful the static framework for classical molecular biology has been, given that organisms are obviously dynamic entities undergoing development over time.

Conclusions: An Invitation

Molecular biology has not received the extent of philosophical attention it deserves, and the little it has received has been limited to the classical period (see Darden and Tabery 2005 for a more detailed summary than what has been presented here).

There are at least two reasons why philosophers should invest more work on the subject:

- Without at least a partial methodological commitment to molecular concepts and techniques, any subdiscipline within biology will likely soon be relegated to irrelevance. Philosophy of biology that does not take molecular biology into account will remain incomplete.
- Modern molecular biology raises fundamentally new epistemological questions, especially about the relevance of physical and semiotic informational accounts that have dominated discussions of biology for the last century. The deployment of philosophical (particularly formal) techniques may contribute significantly to the advancement of the field.

The most important task in the philosophy of biology for the next few decades will be to conceptualize the functional role of DNA within the cell so as to explain the surprising organization and other properties of the genome that were discussed earlier. Physical and informational accounts will probably have to interact in order to create a consistent satisfactory picture. As the last section indicates, any such attempt must necessarily begin with a clearer account than what is currently available of what 'information' means in a biological context. This is probably where philosophers have the most to contribute to the future of molecular biology (see Biological Information). Perhaps techniques from formal epistemology or semantics will enable progress where traditional biological tools have largely failed.

SAHOTRA SARKAR

References

Barrell, B. G., G. M. Air, and C. Hutchison III (1976), "Overlapping Genes in Bacteriophage PhiX174," *Nature* 264: 34–41.

Berget, S., C. Moore, and P. Sharp (1977), "Spliced Segments at the 5′ Terminus of Adenovirus 2 Late mRNA," *Proceedings of the National Academy of Sciences (USA)* 74: 3171–3175.

Berk, A., and P. Sharp (1978), "Structure of the Adenovirus 2 Early mRNAs," *Cell* 14: 695–711.

Cattaneo, R. (1991), "Different Types of Messenger RNA Editing," *Annual Review of Genetics* 25: 71–88.

Cavalier-Smith, T. (1978), "Nuclear Volume Control by Nucleoskeletal DNA, Selection for Cell Volume and Cell Growth Rate, and the Solution of the DNA C-Value Paradox," *Journal of Cell Science* 34: 247–278.

Chow, L., R. Gelinas, T. Broker, and R. Roberts (1977), "An Amazing Sequence Arrangement at the 5′ Ends of Adenovirus 2 Messenger RNA," *Cell* 12: 1–18.

Cook-Deegan, R. (1994), *The Gene Wars*. New York: Norton.

Crick, F. H. C. (1958), "On Protein Synthesis," *Symposia of the Society for Experimental Biology* 12: 138–163.

——— (1973), "Project K: "The Complete Solution of *E. coli*," *Perspectives in Biology and Medicine* 17: 67–70.

Culp, S. (1995), "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles," *Philosophy of Science* 62: 438–458.

Darden, L., and J. Tabery (2005), "Molecular Biology," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2005/entries/molecular-biology

Fox, T. D. (1987), "Natural Variation in the Genetic Code," *Annual Review of Genetics* 21: 67–91.

Gehring, W. (1998), *Master Control Genes in Development and Evolution: The Homeobox Story*. New Haven, CT: Yale University Press.

Gerhart, J., and M. Kirschner (1997), *Cells, Embryos, and Evolution*. Oxford: Blackwell Science.

Griffiths, P. (2001), "Genetic Information: A Metaphor in Search of a Theory," *Philosophy of Science* 67: 26–44.

Hahn, M. W., and G. A. Wray (2002), "The G-Value Paradox," *Evolution and Development* 4: 73–75.

Hull, D. (1972), "Reduction in Genetics—Biology or Philosophy?" *Philosophy of Science* 39: 491–499.

Jacob, F. (1973), *The Logic of Life: A History of Heredity*. New York: Pantheon.

Keller, E. F. (2000), *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Kitcher, P. (1984), "1953 and All That: A Tale of Two Sciences," *Philosophical Review* 93: 335–373.

Koshland, D. E., Jr. and K. Hamadani (2002), "Proteomics and Models for Enzyme Cooperativity," *Journal of Biological Chemistry* 277: 46841–46844.

Lercher, M. J., A. O. Urrutia, and L. D. Hurst (2002), "Clustering of Housekeeping Genes Provides a Unified Model of Gene Order in the Human Genome," *Nature Genetics* 31: 180–183.

Lolle, S. J., J. L. Victor, J. M. Young, and R. H. Pruitt (2005), "Genome-Wide Non-Mendelian Inheritance of Extra-Genomic Information in *Arabidopsis*," *Nature* 434: 505–509.

Mattick, J. (2003), "Challenging the Dogma: The Hidden Layer of Non-Protein-Coding RNAs in Complex Organisms," *BioEssays* 25: 930–939.

Mattick, J., and M. J. Gagen (2005), "Accelerating Networks," *Science* 307: 856–857.

Mayr, E. (1961), "Cause and Effect in Biology," *Science* 134: 1501–1506.

Monod, J. (1971), *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. New York: Knopf.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Newman, M. E. J. (2003), "The Structure and Function of Complex Networks," *SIAM Review* 45: 167–256.

Olby, R. C. (1974), *The Path to the Double Helix*. Seattle: University of Washington Press.

Pauling, L., and R. B. Corey (1950), "Two Hydrogen-Bonded Spiral Configurations of the Polypeptide Chains," *Journal of the American Chemical Society* 71: 5349.

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

Sarkar, Sahotra (1989), "Reductionism and Molecular Biology: A Reappraisal," Ph. D. Dissertation, Department of Philosophy, University of Chicago.

——— (1996), "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology," in Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht, Netherlands: Kluwer, 187–231.

——— (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

——— (2005), *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge, MA: MIT Press.

Schaffner, K. F. (1967), "Approaches to Reduction," *Philosophy of Science* 34: 137–147.

Shannon, C. E. (1948), "A Mathematical Theory of Information," *Bell System Technical Journal* 27: 379–423, 623–656.

Stephens, C. (1998), "Bacterial Sporulation: A Question of Commitment?" *Current Biology* 8: R45–R48.

Tauber, A. I., and S. Sarkar (1992), "The Human Genome Project: Has Blind Reductionism Gone Too Far?" *Perspectives on Biology and Medicine* 35: 220–235.

Thiéffry, D., and S. Sarkar (1998), "Forty Years under the Central Dogma," *Trends in Biochemical Sciences* 32: 312–316.

Waddington, C. H. (1962), *New Patterns in Genetics and Development*. New York: Columbia University Press.

Waters, C. K. (1990), "Why the Anti-Reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics," in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990: Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association*. East Lansing, MI: Philosophy of Science Association, 125–139.

Watson, J. D., and F. H. Crick (1953), "Molecular Structure of Nucleic Acids—A Structure for Deoxyribose Nucleic Acid," *Nature* 171: 737–738.

Williams, K. L., and D. F. Hochstrasser (1997), "Introduction to the Proteome," in M. R. Wilkins, K. L. Williams, R. D. Appel, and D. F. Hochstrasser (eds.), *Proteome Research: New Frontiers in Functional Genomics*. Berlin: Springer, 1–12.

Wimsatt, W. C. (1976), "Reductive Explanations: A Functional Account," *Boston Studies in the Philosophy of Science* 32: 671–710.

Yockey, H. P. (1992), *Information Theory and Molecular Biology*. Cambridge, UK: Cambridge University Press.

See also **Biological Information; Biology; Explanation; Function; Genetics; Heredity and Heritability; Holism; Mechanisms; Neurobiology, Philosophy of; Genetics; Physicalism; Reductionism**