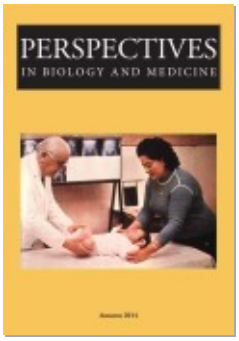
The Human Genome Project: Has Blind Reductionism Gone too Far?

Alfred I. Tauber, Sahotra Sarkar



Perspectives in Biology and Medicine, Volume 35, Number 2, Winter 1992, pp. 220-235 (Article)

Published by Johns Hopkins University Press
DOI: 10.1353/pbm.1992.0015

THE HUMAN GENOME PROJECT: HAS BLIND REDUCTIONISM GONE TOO FAR?

*ALFRED I. TAUBER and SAHOTRA SARKAR**

Introduction

The Human Genome Project (HGP) is now well under way. An internationally coordinated effort, its ultimate aim is to sequence the entire human genome, producing a string of As, Ts, Gs, and Cs that would fill 13 sets of *Encylopaedia Britannica* [1]. This is the largest and most expensive biology project ever proposed, let alone undertaken. In the United States, where most of the effort has so far been concentrated, the program has been split between the Department of Energy (DOE), whose original interest arose from the possible utility of direct DNA analysis to detect radiation-induced mutations, and the National Institutes of Health (NIH) [2]. From its first conception, the feasibility and advisability of the project have been vigorously debated. After 3 years of discussion, an almost universal consensus has emerged in favor of a comprehensive mapping program to locate genes precisely on chromosomes (as opposed to total sequencing). This is expected to have immediate benefits for molecular studies of human genetic disease, especially for diagnosis (prenatal, preclinical, carrier), of single-locus, multifactorial, and chromosome disorders [3, 4]. Most would not argue the merits of that issue, but the question of complete sequencing is still to be resolved, since not all agree that such a goal "will yield a harvest of information that will drive the research enterprise for at least the next 100 years" [5].

The purpose of this article is, first, to evaluate critically the scientific merit of the project; second, to very briefly suggest that the project is the culmination of a research program within biology that, for 2 centu-

This article is dedicated to the memory of Salvador Luria. It is contribution BTBG 91-1 from the Boston Theoretical Biology Group, Boston University.

*Departments of Philosophy, Pathology, and Medicine, Boston University, 80 East Concord Street, Boston, Massachusetts 02118-2394.

0031-5982/92/3502-0776$01.00

ries, has been fundamentally reductionist; and, third, to conclude that the doubts raised about the scientific merits of the project may well illustrate the limitations of such reductionist approaches in biology (and, perhaps even more strongly, in medicine) at present. This is not an exhaustive review of the history of HGP or of all the controversies surrounding its purpose, economics, or politics. These are already receiving systematic treatment [6–8] and are considered here only when directly relevant to this article's immediate aims.

Fact, Fiction, and Fancy in the Scientific Aims of HGP

The main proponents of HGP have come from the top echelons of the molecular biology establishment. Some have been nothing short of elegiac about the project. For instance, James Watson has argued that sequencing the human genome is analogous to placing a man on the moon [9]. For such strident public advocacy, Watson has been rewarded with the czarship of the program, as it emerged as a somewhat coherent whole after fierce federal interdepartmental fighting between mid-1985, when it was first articulated as a realizable public objective [10], and late 1988, when some advisory committees were formed at the NIH [11]. Initially, the DOE and NIH vied for control of the project, but they eventually reached a compromise, encapsulated in an unorthodox "Memorandum of Understanding" [12] that explicitly recognized the DOE's interest in what, most naturally, should have been an NIH or NSF project.

More important than such organizational compromises, by 1989 a consensus about scientific strategy had also emerged among the HGP's proponents. This consensus, presented as a report to Congress in January 1990 [13], was that sequencing the human genome was a worthwhile end, but that before the entire genome was actually sequenced, the next 5 years should be devoted to the construction of genetic linkage maps, physical maps, and DNA sequencing technique development. The linkage map is expected to identify and locate genes that are now recognized only by their phenotypic effects. In addition, assembly of the physical map appears to require the linkage studies, although the strategic order of these studies is still in debate as the mapping techniques continue to evolve [14].

Though the proponents of HGP have achieved consensus, the rest of the community, scientific and lay, have not. Prominent scientific figures such as Bernard Davis [15] and Salvador Luria [16] remained more than skeptical, while historians, philosophers, and sociologists of science have been even more worried [6, 7, 17]. These concerns are in two broad categories. First, there have been persistent questions about the project on the basis of its scientific and medical merits (i.e., about the utility of

the research program, including its characterization of its aims); related questions concerning the economic influence on divergent and related biomedical research; and perhaps more nebulous questions based ultimately on sociological considerations such as the likely effect of "big science" on the creativity of the individual investigator-oriented research that has hitherto been predominant in biology [18]. Second, significant concerns have emerged about the ethical, political, and social ramifications of the project [19, 20]. Ethicists and policy analysts have worried about potential eugenic implications, the unpreparedness of the medical community to provide the adequate genetic counseling that will be in high demand [6], the likely use of sequence information by employers and insurers [6], and the possibility of the creation of a "biological underclass" in a political economy in which the prospects for socialized medicine are dim in the immediate future [7].

This article will not address any of the questions from the second category. This is not to deny, in any sense, that they are important. However, those debates have already begun in earnest, and to the credit of the proponents of HGP, significant funding is available for studies of the ethical, political, and social adumbrations of the project. At NIH, for instance, 3 percent of the entire budget for the project has been earmarked for such studies. The focus here, however, will be on the scientific merits of the project, which have not been receiving similar scrutiny, especially now that the project has moved ahead. If the considerations presented here have any merit, the scientific relevance of the project is even more doubtful than has previously been suggested.

The scientific aim of HGP is the entire sequence of the human genome, and, if all works out as anticipated, at the end this sequence will be known. Any analysis of HGP must, therefore, consider the utility and importance of realizing this aim. Some of the chief proponents of the project, like Watson and Gilbert, have argued solely on scientific grounds. According to Gilbert [21], once the sequence of any organism is available, its biology can emerge from being a purely experimental field to a theory-based science. The sequence can be stored in a computer database and the entire corresponding amino acid sequences of proteins read off, and theoretical analyses can begin. In principle, every control mechanism can be obtained from this information, thereby answering a fundamental question of biology. Next, the interaction of genes can also be obtained, and, finally, the entire behavior of the organism calculated.

Most doubts about the scientific importance of HGP ultimately emerge from the observation that these scientific expectations are at best naive: they ignore the complexity of the issues and the limitations that it can impose on the possibility of such theoretical computation.

One source of difficulty is the protein-folding problem [22]. For more than 3 decades now, it has been assumed that the primary structure or amino acid sequence of a protein determines its tertiary structure or three-dimensional conformation. In principle, therefore, one should be able to calculate the conformation from the sequence information. However, after decades of systematic, and sometimes quite sophisticated, computational efforts, the protein-folding problem is yet to be solved. Even for as simple a protein as bovine pancreatic trypsin inhibitor, which is a mere 58 residues long, the conformation cannot be predicted from the sequence. For other functional proteins that can be several thousand residues long, the prospects for such prediction are bleak.

Yet, amino acid sequences are the most that can be directly read off from the sequence of the genome. Even such direct translation is not straightforward. Large segments of the genome (perhaps as many as 90 percent) do not code for anything at all. Presumably the boundaries between such sequences and coding sequences consist of some specific short patterns (the "intron [noncoding]—exon [coding]" borders). But not all of these are known. Thus, even with the complete DNA sequence of the genome at hand, it would still not be possible to state with any confidence exactly what amino acid sequences end up being expressed.

Further, the current insolubility of the folding problem means that even when amino acid sequences are known, the three-dimensional conformation of those proteins cannot be theoretically computed. Yet, the current understanding of molecular biology prescribes that it is structure that determines function of proteins. A solution to this problem does not seem immediately forthcoming. There are, however, two potential ways out; but even these do not appear to be practical. One is to obtain the three-dimensional conformation of a protein directly by crystallizing it and experimentally solving the crystal. But this is a laborious process. As a result, comparatively few proteins have already been solved. It is not even remotely likely that enough of the thousands of proteins in the body will soon have their structure thus discovered so as to permit the sort of theorizing envisioned by Gilbert.

The other potential solution is to note that only the structure of a small part of a protein, the "active site," is usually involved in its function. But even characterizing the active site cannot be accomplished theoretically unless the major features of the rest of the protein can somehow be determined. Consequently, unless the folding problem is solved as a first step, there is no practicable way to begin characterizing biological function at the molecular level from sequence information alone. If theoretical biology is the aim, it would therefore seem more rational, in such a situation, to pool major resources for all problems of biological function at the molecular level. Moreover, at the physiological

level, function can be understood without precise delineation of protein structure. The appropriate level can be determined only from the biological context.

The folding problem is emphasized here only because of the critical role that proteins play in mediating virtually all biological interactions and because the primary structure of proteins is mostly what the DNA sequence codes for. Other relevant problems whose solution is critical for understanding biological function include, to list just a few, the epigenetic control of DNA transcription and related problems of developmental and regenerative phenomena, organizational explanation of complex systems such as the immune and neuronal networks, programmed cell death, and many other poorly understood temporally sequenced events, or even activation of proto-oncogene function. These are but examples of complex biological function and behavior that may ultimately utilize DNA sequence data but whose systematic exploration must first exploit other experimental strategies. The design of complex integrative models of these phenomena cannot rely on the HGP to offer solutions to questions arising at these other levels of biological organization. Long before a relentless, but blind, quest for DNA sequences was begun, the potential utility of such information relative to that of other experimental data should have been much more clearly explicated.

Moreover, even if such a primary issue as the protein-folding problem is resolved, the prospects for predictive theorizing about biological behavior still remain dim. Even if all control mechanisms can be obtained from sequence information—and it is difficult enough to imagine how this would be the case—what matters most in typical biological processes, such as development, is the time of action of genes [23]. This information is at least partly in the environment, not in the genome. Further, to what extent it is in the genome is still unknown. Yet, for theorizing, this information is crucial. Once again, it seems more rational to pool limited resources on something other than the blind quest for all DNA sequences. In this case, if theoretical understanding is truly the object, studies of gene regulation during development seem more promising than the search for sequences. After all, as already mentioned, as much as 95 percent of the genome may not be involved in any coding or control anyway [24].

To compound all, the human genome has on the order of 100,00 genes. That the interactions of these genes can be theoretically understood from sequence information is wildly implausible. Much more likely, these interactions will have to be studied conventionally, using phenotypic properties: knowing the sequence of all but the relevant genes in a few coding regions is largely irrelevant to this process.

So much for Gilbert's type of theorizing. But perhaps it can be suggested that the sequence will be useful for evolutionary biology, if not

for the type of predictive program that Gilbert and others envision. This is true only to a very limited extent and hardly justifies the concentration of human and economic resources on the HGP. Evolutionary biology thrives on comparisons, and, as James F. Crow has often emphasized, it would be better to know 10 percent of the genomes of 10 different species than the entire genome of any single one.

As originally conceived, the HGP deifies the human genome, and only recently has serious consideration been given to some other species, including *E. coli,* yeast, *Drosophila,* and mouse. These other species, however, are being studied primarily to hone techniques for human genome sequencing, not for their intrinsic interest. There is little doubt that evolutionary biology could benefit somewhat from the knowledge of these sequences. After the completion of the HGP, families of related genes will be able to be sequenced better than at present because of the new techniques developed. To the extent that evolutionary biology benefits from the HGP, it is only from such fallout, and similar accomplishments could have been achieved for a mere fraction of the resources designated for the HGP.

Moreover, for evolutionary studies, linkage maps and an understanding of epistasis (or gene interaction) and gene regulation are of critical importance at present. Sequences, although now being used as another type of information in systematics to construct phylogenies, have yet to yield much insight. For theoretical biology, too, as has been noted, the understanding of these features and the functional characterization of biological structures at all levels of organization are currently most needed. Presumably with such reservations in mind, and because so much of the DNA in the genome does not seem to code for anything, some, like Brenner [24] and Weinberg [25], have argued for sequencing, but only on a very limited basis.

Weinberg has argued for a strategy directed only at information-containing (coding or regulatory) regions of the genome as found in cDNA libraries made from processed transcripts. However, even this limited sequencing is suspect. As Weis [26] has argued, from the point of view of an experimentalist, when one is cloning a coding sequence of a target protein, already knowing the genomic sequence containing the genes saves relatively little effort. The most laborious steps are the production and screening of cDNA libraries and the subsequent sequencing of the clones, whether or not their expected sequence was already known. Available genome sequences would, therefore, only confirm that effort. Further, intron-exon borders could be determined far more cheaply than by sequencing entire segments of the genome containing the relevant gene.

It is possible, of course, as Berg [27] and McKusick [1] (among others) have pointed out, that the noncoding and nonregulatory region of the

genome will actually turn out to be of considerable interest and that it is simply a prejudice to suggest that it is "junk." In particular, it is possible that the sequence of these regions, if ultimately obtained for many different species, may reveal data of some value to evolutionary biology. Similarly, they may reveal information about higher levels of biological control and regulation than currently known. However, the mere existence of such possibilities does not justify blind sequencing. If these possibilities are remotely probable, it is incumbent for the proponents of HGP to so argue their case—which, as Davis [15] has pointed out, they have not done. Moreover, it is far more probable from what is already known that quite a significant amount of the information for the control of genes is present as epigenetic information, such as methylation patterns of DNA that are inherited during replication. The design of the HGP simply ignores the relevance of this possibility.

There is yet another conceptual problem with the HGP, and its effects pervade all the scientific and medical fields for which the HGP's advocates promise the most benefit. If the problems discussed above illustrate a reductionist naïveté about biochemistry and organismal biology, this one exemplies an even more egregious naïveté about the nature of evolution; at the end of the HGP, there will be available a single sequence that is ostensibly that of the "human genome." This may well become a measure of the "normal" genome. But whose genome will this be, Watson's?

The point is that, as many critics have indicated, there simply is no single entity as the human genome. The amount of variability in any natural population is immense, and it is this variability that characterizes the object of study of evolutionary biology. In human populations this variation can be expressed in terms of polymorphism at both the DNA and the protein levels.

Genetic polymorphism at the protein level has been known since Landsteiner described the ABO blood groups in 1901, and Garrod [28] ended his famous 1902 paper on alkaptonuria with a prophetic statement of the eventual understanding of the genetic basis of disease: "just as no two individuals of a species are absolutely identical in bodily structure neither are their chemical processes carried out on exactly the same lines." There is obviously less observed polymorphism at the protein level than actually subsists at the DNA level because of the degeneracy of the code. But even at the protein level, the degeneracy is enormous. By electrophoretic analysis of proteins, 28 percent of loci show polymorphism, and the average heterozygosity per locus is approximately 20 pecent [29].

Consider hemoglobin, at present the best characterized human protein [30]. It has been demonstrated that there are at least 450 variants of hemoglobin, of which almost half are fully functional and, therefore,

physiologically silent. This amount is probably a serious underestimate. Of 2,583 possible base substitutions of the DNA coding for the alpha- and beta-chains of hemoglobin, only 1,690 would result in amino acid substitutions because of the degeneracy of the code. Of these, only 575 (or about 45 percent) would result in a change that could be detected by electrophoresis, and that is what has largely led to the discovery of the known 450 variants. The true amount of protein polymorphism could thus be much higher. Among the fully functional variants, it is fallacious and even dangerous to call any one "normal," simply because any notion of normality that goes beyond functionality has no scientific, medical, or philosophical basis [17] and because such unjustified attributions of normality may well provide grounds for discrimination [6, 7].

Moreover, DNA polymorphism, the relevant concept for defining heterogeneity at the genomic level, vastly exceeds what can be observed through protein studies. Restriction enzymes cleave DNA at specific sites, and different pieces may be identified by their mobility after radioactive labeling. These so-called restriction fragment length polymorphisms (RFLPs) have revealed that the extent of nucleotide diversity is about an order of magnitude greater than that observed from data for structural genes that specify gene products. On the average, one of 200–500 nucleotides will differ between random chromosomes [29]. In addition, there is variability of tandem repeat sequences, which leads to further complexity (deletions and duplications by unequal crossovers) designated as hypervariable regions. With such power to detect individuality, the RFLP profle is (controversially) being used even for forensic purposes. It is almost certainly a signature of unique genetic constitution.

Ironically, the use of RFLPs to construct the human genetic linkage map [31] is based on the existence of this enormous genomic polymorphism, and yet this is what some major proponents of the human genome project attempt to denigrate. Consider McCusick's response to critics:

> Such variation can be overemphasized in the human-genome initiative, however. The question often asked, especially by journalists, is "Whose genome will be sequenced?" The answer is that it need not, and surely will not, be the genome of any one person. Keeping track of the origin of the DNA that is studied will be important, but the DNA can come from different persons chosen for study of particular parts of the genome. Such an approach is consistent with that of most biological research, which depends on a few, and even on single individuals, to represent the whole, and with the fact, well recognized by geneticists, that there is no single normal, ideal, or perfect genome. [1, p. 913]

Moreover, Gilbert [18] has even emphasized that the nonvariable parts of the genome reflect the "commonality" of the human species.

Although such arguments do address the somewhat naive objections

of those who indeed seriously thought that the genome of some particular individual would end up as the genome being sequenced, they do not seem even to understand, let alone address, the precise nature of the variability involved. No matter where each segment of the genome came from, it is still one segment, and, from everything that is currently known, the population thrives on variations of it. There is no comprehensible way in which it can "represent" the "genome of the species."

The following contrast drives this point home. The pictures of the human body in any anatomy textbook claim to represent the human body in general. Every actual body deviates from these pictures. The sense in which these pictures represent "the human body" is that certain topological relations between body parts and certain spatial relations between these parts are maintained. This is good enough for such representations, since (1) these relations cannot easily be violated in ordinarily functioning bodies, and (2) no part of these representations corresponds to any particular body.

On both counts, any analogy to the human genome breaks down. The variability is much greater in any natural population, and, in any sequence finally reported, some particular individual's DNA bit will have been sequenced. Further, and more important, the sequence consists of highly irregular information; there is simply no known sense in which such information can ever be "represented" on the average. Indeed, a possible analogy to the anatomical case is a linkage map that shows where alleles reside, and, while justifying mapping, sequencing does not necessarily follow. Moreover, even this analogy is incomplete: recombination and other similar factors routinely change linkage relationships. This is an important source of human (and other intraspecific) variability and a major cause of the difficulty—and beauty—of theoretical population genetics.

Reductionism in Molecular Biology

When a scientific research program goes awry, becomes moribund, or simply becomes as mechanical as the blind quest for DNA sequences, it is always pertinent to look back and analyze how it arrived at that stage. The HGP is the ultimate product of an extreme reductionist vision of biology that has held that *to understand better one need only to go smaller.* Reductionism in molecular biology constitutes a research program that attempts to explain and understand biological systems completely in terms of the physical interactions of their parts [32]. From that point of view, it is natural to assume that fundamental understanding of biology comes only from the level of DNA, the alleged blueprint for living systems. Indeed, this is exactly what Gilbert and Watson have

routinely suggested, but as the discussion in the last section demonstrates, this vision has by now become quite limited.

The reductionist program of contemporary molecular biology has its roots in the seventeenth-century mechanical philosophy according to which all phenomena were to be explained in terms of the motions of the constituent particles of bodies and contact interactions between them. The contemporary reductionist is more sophisticated and does not insist on only contact interactions but is willing to admit any physical or chemical mechanism that is warranted. Indeed, this relaxed attitude toward permissible interactions goes back to Helmholtz, who, in the mid-nineteenth century, presented an articulate relaxation of the mechanical philosophy to allow central forces such as gravitation and the electrostatic force [33].

It is to Helmholtz and his like-minded colleagues that the lineage of modern reductionism in biology can be most directly traced back. In the 1840s they embarked on an ambitious program to purge biology of any traces of vitalism by providing quantitative mechanical accounts of biological phenomena [34, 35]. Using a careful quantitative technique, Helmholtz, for example, measured the heat generated by contracting muscles, and showed that conversion and conservation of energy held for this phenomenon exactly as they did in the inanimate world. This was generalized into a belief that exactly all, and only, the laws of inorganic physics and chemistry were operative in the organic world. A complete demonstration of this assumption was obviously not forthcoming, given the complexity of organic phenomena and the state of experimental techniques. Passing beyond empirical demonstration alone, therefore, they transformed their belief into a metaphysical faith in reductionism—which was fecund in producing experimental programs.

This reductionist position was not unchallenged. A sophisticated group of embryologists and physiologists centered around von Baer attempted to chart a different explanatory program in biology that avoided both such complete reductionism and the vitalism of the eighteenth century, which had already been discredited. Their emphasis was on the unity of an organism, on the purposive nature of its behavior, and on the functions of its parts. These parts were structured according to some organizing principle that conferred integrity and unity to the organism as a whole [35].

The debates between these two programs were fierce. Arguably, the reductionists prevailed in the late nineteenth century as the new discipline of biochemistry emerged with a strong accent on the quantitative study of enzymes [36]. The dominating figure was Hopkins at Cambridge, whose faith in the mechanical explanation of biological phenomena was unbounded. The roots of molecular biology lie in both biochem-

istry and the related discipline of structural chemistry, but, perhaps unexpectedly, also in the type of thinking associated with their nineteenth-century critics.

In the latter regard, the vital role was played by the founder of the quantum theory, Niels Bohr, the son of physiologist Christian Bohr, who, along with J. S. Haldane, was also noted for putting explanatory emphasis on function over mechanism in the study of physiological responses such as that of hemoglobin to oxygen and carbon dioxide. Christian Bohr and Haldane both acknowledged that mechanical accounts of biological phenomena were often possible. However, implicit in their work are the assumptions (1) that their functional accounts were, in some sense, "independent" of these; (2) that they are more valuable for biological understanding; and (3) that mechanical accounts of biological phenomena could never be complete. This dualistic approach was taken up by Niels Bohr. In quantum mechanics it led to his controversial principle of complementarity whereby, in the quantum description of nature, conflicting modes of interpretation, such as those in terms of particles or waves, would be necessary because no single mode could ever be complete. More important, in biology, it led to Bohr's conclusion that the phenomenon of life could never be explained solely on the basis of the laws of physics.

Bohr's views, first publicly expounded in 1932, were instrumental in the establishment of molecular biology [37]. They inspired Max Delbrück, then a young physicist, to abandon physics and take up biology with the hope of discovering exactly where physical explanation of biological phenomena collapsed. Delbrück began, for the lack of any cogent alternative, simply by attempting such physical explanation as systematically as possible. It immediately led, for example, to a quantum mechanical theory of the gene that was later popularized by Schrödinger [38] in *What Is Life?* This constructive aspect of Delbrück's program, especially after the appearance of Schrödinger's book, drew Salvador Luria, Maurice Wilkins, Watson, and many others into biology, although the romantic dream of the limitations of physics attracted Seymour Benzer and Gunther Stent.

Meanwhile, Delbrück's interest had focused on phage after Stanley crystallized the first virus [39]. This important development suggested to Delbrück that viruses, while living, were still simple and small enough to be studied by the methods of physics. If the laws of physics were to prove insufficient to explain the phenomena characteristic of life, such as gene replication, presumably such failures could most easily be discovered in viruses. Out of Delbrück's interest and that of Luria and Hershey emerged the Phage Group, which, through training and proselytization, introduced most of the techniques that were critical for the early development of molecular biology. In particular, the work of Her-

shey and Chase verified Avery's earlier observation that the hereditary material was DNA, not protein as had long been almost universally assumed.

Not all members of the Phage Group, however, had much sympathy for Bohr's romantic views. Hershey, for one, had little time for such "complementarity double-talk" [40]. No exception to the laws of physics ever emerged from the work of the Phage Group. Finally, the formulation of the double helix model for DNA in 1953 [41] destroyed virtually all hope that any such would be forthcoming. The double helix model showed explicitly how genetic information can be stored in sequences of nucleotide bases, how such sequence information can be mechanically transferred during cell replication. The model was as reductionist as possible. As Delbrück later observed:

> Indeed we might say that the discovery of the Double Helix in 1952 did for biology what many physicists had longed for in atomic physics: a resolution of all the miracles in terms of classical mechanical models, not requiring an abdication of our customary intuitive expectations. The Double Helix, indeed! With one blow the mystery of gene replication was revealed as a ludicrously simple trick, making those who had expected a deep solution feel as silly as one might feel when shown the embarrassingly simple solution to a chess problem one may have struggled with in vain for a long time. [42]

Thus collapsed an explicitly antireductionist research tradition that was critical to the early development of molecular biology.

Since then, research in molecular biology has been almost universally reductionist as it has probed deeper and deeper into biological phenomena. The general strategy of research (almost entirely experimental) has been toward the explanation of properties of biological systems in terms of the interactions of their constituent parts. At its best, it has chosen problems of function that were already perceived to be of biological interest. Thus came the operon model of gene regulation [43] and, less successfully, the speculative theories of the nature of the genetic code [44]. At its worst, research has been driven by investigative programs not sufficiently cognizant of any broader biological contexts. The blind sequencing of DNA is just such a case. Thus has come about the Human Genome Project.

Conclusion

This article has systematically emphasized the scientific limitations of HGP and argued that the program has emerged from a radically reductionist research tradition in molecular biology. From these two conclusions, an intriguing insight emerges: that there are important limits to blind reductionism in molecular biology and, perhaps more important,

to its use in medicine. For molecular biology, the following conclusions may be drawn:

1. Going to the lowest level of organization may not yield any insight of interest. Many of the human (and other) genomes consist of noncoding stretches of DNA. Sequencing these may well be irrelevant to understanding biological function; and it is only barely possible it will yield any new evolutionary insight even as a fallout. Insight may emerge from this strategy only if something unexpected emerges, but it is hardly justifiable in any science to look at random for the unexpected.

2. Reductionist explanation, even when possible, is not cost-effective in effort expended [45]. An automobile design engineer must worry about thermodynamic considerations. No doubt every conclusion drawn from thermodynamic principles can be obtained, at least approximately, from the kinetic theory, but no sensible engineer would use the kinetic theory to decide the specifications for a car engine. In the case of HGP, the relevant higher level is that of the gene as opposed to the level of sequence. Mapping is justified; blind sequencing is not.

3. The sheer complexity of a system may make reductionist explanation impossible. Although it is traditional among philosophers of science to emphasize the "in principle" possibility of reductionist explanations in such cases, such claims are nothing but an assertion of confidence in the principles of the reducing theory. They do not provide actual explanations—which, after all, is what science is about. The lessons learned from the intractability of the protein-folding problem, when extended to HGP, show that the knowledge of the sequence will not necessarily permit explanation of phenomena of interest. Whether such explanation is in principle possible is irrelevant.

For medicine, moreover, some additional lessons can be drawn; and these, too, must be considered because HGP has been justified by some of its most prominent proponents in terms of potential medical benefits.

1. The complexity of understanding most biological phenomena by starting purely at the level of DNA sequences, as outlined above, carries over into medicine. Even in the case of single-locus diseases, effective intervention is by no means certain when the sequence is known, as decades of experience with sickle cell disease has shown [30]. Furthermore, to the extent that genetic factors are important in the etiology of various diseases, current evidence suggests that several loci are almost always simultaneously involved [46]. This complicates matters in exactly the same way as in the prediction of biological behavior or function from DNA sequence information. It may be suggested that new techniques of gene therapy would resolve all these problems, but that utopian promise only addresses relatively few diseases identified as due to single protein defects. For complex disease processes involving poorly understood

multiple factors or disease susceptibility, before gene therapy can be responsibly suggested all the relevant epistatic interactions and pleiotropic effects of the participating genes must be characterized. This, as in all the other instances discussed here, involves investigation at levels higher than the DNA sequence. And, finally, when some knowledge of the DNA sequence is necessary, only the sequence of the coding regions of the genes, that is, a tiny fraction of the genome, and certainly not the entire sequence, becomes relevant.

2. Many of the genes implicated in disease have low or highly variable expressivity [29], including the gene for the sickle cell trait. In such circumstances, an extragenetic factor is critical to the etiology of the disease. There is no reason to believe that such an environmental component is in general due to the epistatic effects of genes at other loci. Sequence information is of little value in these cases, since what has to be understood is the interactions with the environment that, at least to a first approximation, consists of or is mediated by biological entities at higher levels of organization than the DNA sequence. Once again the quest for DNA sequences offers little hope for adequate therapeutic intervention.

This article has emphasized the scientific shortcomings of the HGP. It is important to note, as many have pointed out, that the original design of the HGP has been extensively modified as a result of criticism and concerns over the scientific legitimacy of its initial goals. The only consensus that has survived is a limited acquiescence to the idea of mapping. That consensus has not been challenged here. Rather, the criticisms offered are of the original and ultimate aims of the HGP: namely, blind sequencing.

In conclusion, it is important to reemphasize that biological systems are complex, largely hierachically organized, interacting systems; to understand them, the most critical investigation is that of the function of units at each level of organization. This has largely been the purpose of biological research; in even as reductionist a science as molecular biology, few would argue with Davis: "Our fundamental goal is to understand the human genome and its products, and not to sequence the genome because it is there" [15]. After the initial reductionist euphoria of the 1950s and early 1960s, molecular biology has often paid attention to such problems in spite of generally being reductionist: hence the interest in studies of molecular evolution, of development at the molecular level, and of molecular mechanisms that mediate the neural or immunological systems. But these, like the other examples mentioned above, are problems independently known to be of biological importance. In many of these areas, progress has been unexpectedly slow, but perhaps the lesson to be learned is not the desirability of more reduction

but less—the lack of progress may well be due to the limits of blind reductionism itself. If these considerations have any merit, the blind sequencing of the HGP is an unfortunate step backward.

REFERENCES

1. McKusick, V. A. Mapping and sequencing the human genome. *N. Engl. J. Med.* 230:910–915, 1989.
2. McKusick, V. A. The human genome organization: history, purposes, and membership. *Genomics* 5:385–387, 1989.
3. Rossiter, B. J. F., and Caskey, C. T. Molecular studies of human genetic disease. *FASEB J.* 5:21–27, 1991.
4. McKusick, V. A. Current trends in mapping human genes. *FASEB J.* 5:12–20, 1991.
5. Collins, F. S. The genome project and human health. *FASEB J.* 5:77, 1991.
6. Holtzmann, N. A. *Proceed with Caution.* Baltimore: The Johns Hopkins Press, 1989.
7. Nelkin, D., and Tancredi, L. *Dangerous Diagnostics: The Social Power of Biological Information.* New York: Basic Books, 1989.
8. Davis, J. *Mapping the Code: The Human Genome Project and the Choices of Modern Science.* New York: Wiley, 1990.
9. Watson, J. D. The human genome project: past, present and future. *Science* 248:44–49, 1990.
10. Sinsheimer, R. L. The Santa Cruz Workshop—May, 1985. *Genomics* 5:954–956, 1989.
11. Roberts, L. Genome project underway, at last. *Science* 243:167–168, 1989.
12. Cantor, C. R. Orchestrating the human genome project. *Science* 248:49–51, 1990.
13. DHHS. *A Five-Year Plan for the Human Genome Project.* Washington, D.C.: DOE, 1990.
14. National Research Council Committee on Mapping and Sequencing the Human Genome. *Mapping and Sequencing the Human Genome.* Washington, D.C.: National Academy Press, 1988.
15. Davis, B. D. The human genome and other initiatives. *Science* 4:2941–2942, 1990.
16. Luria, S. E. Human genome project. *Science* 243:878, 1989.
17. Sarkar, S. Structures of choice in human genetics. Paper presented at the meeting on Historical and Social Study of the Human Genome Initiative, Department of the History of Science, Harvard University, Cambridge, June 15, 1990.
18. U.S. Congress. *Mapping Our Genes—Genome Projects: How Big? How Fast?* Office of Technology Assessment, OTA-BA-373, Washington, D.C.: GPO, 1988.
19. Anderson, W. F. Human gene therapy: scientific and ethical considerations. *J. Med. Philos.* 10:275–291, 1985.
20. Position paper on human genome initiative. Committee for Responsible Genetics, Boston, 1990.
21. Gilbert, W. Current state of the HGI. Paper presented at the meeting on Historical and Social Study of the Human Genome Initiative, Department of the History of Science, Harvard University, Cambridge, June 15, 1990.
22. Gierasch, L. M., and King, J. (eds.). *Protein Folding: Deciphering the Second*

Half of the Genetic Code. Washington, DC: American Association for the Advancement of Science, 1990.
23. DAVIDSON, E. H. *Gene Activity in Early Development,* 3d ed. New York: Academic Press, 1986.
24. BRENNER, S. The human genome: The nature of the enterprise. In *Human Genetic Information: Science, Law and Ethics,* edited by D. CHADWICK. Chichester: Wiley, 1990.
25. WEINBERG, R. A. There are two large questions. *FASEB J.* 5:78, 1991.
26. WEIS, J. H. Usefulness of the human genome project. *Science* 248:1595, 1990.
27. BERG, P. All our collective ingenuity will be needed. *FASEB J.* 5:75–77, 1991.
28. GARROD, A. E. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 2:1216–1620, 1902.
29. VOGEL, F., and MOTUSKY, A. G. *Human Genetics: Problems and Approaches,* 2d ed. Berlin: Springer, 1986.
30. BUNN, H. F., and FORGET, B. F. *Hemoglobin: Molecular, Genetic and Clinical Aspects.* Philadelphia: Saunders, 1986.
31. SCHUMM, J. W.; KNOWLTON, R. G.; BRAMAN, J. C.; et al. Identification of more than 500 RFLP's by screening random genomic clones. *Am. J. Hum. Genet.* 42:143–159, 1988.
32. SARKAR, S. Reductionism and molecular biology: a reappraisal. Ph.D. dissertation, University of Chicago, 1989.
33. STEIN, H. Some philosophical aspects of natural science. Ph.D. dissertation, University of Chicago, 1958.
34. GALATY, D. H. The philosophical basis of mid-nineteeth century German reductionism. *J. Hist. Med. Allied Sci.* 29:295–316, 1974.
35. LENOIR, T. *The Strategy of Life: Teleology and Mechanism in Nineteenth Century German Biology.* Dordrecht: Reidel, 1982.
36. KOHLER, R. E. The enzyme theory and biochemistry. *Isis* 64:181–196, 1973.
37. BOHR, N. Light and life. *Nature* 131:421–423, 457–459, 1933.
38. SCHRÖDINGER, E. *What Is Life?* Cambridge: Cambridge Univ. Press, 1944.
39. FISCHER, E. P., and LIPSON, C. *Thinking about Science: Max Delbrück and the Origins of Molecular Biology.* New York: Norton, 1988.
40. STENT, G. S. That was the molecular biology that was. *Science* 160:390–395, 1968.
41. WATSON, J. D., and CRICK, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737–738, 1953.
42. DELBRÜCK, M. Light and life III. *Carlsberg Res. Commun.* 41:299–309, 1976.
43. JACOB, F., and MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318–356, 1961.
44. CRICK, F. H. C.; GRIFFITH, J. S.; and ORGEL, L. E. Codes without commas. *Proc. Nat. Acad. Sci. USA* 43:416–421, 1957.
45. WIMSATT, W. C. Reductive explanation: a functional account: *Boston Stud. Philos. Sci.* 32:671–710, 1976.
46. WILLIAMSON, R., and KESSLING, A. M. The problem of polygenic disease. In *Human Genetic Information: Science, Law and Ethics,* edited by D. C. CHADWICK. Chichester: Wiley, 1990.