

JAMES D. WATSON

## A Personal View of the Project

# 7

When I was going into science, people were concerned with questions of where we came from. Some people gave mystical answers—for example, “the truth came from revelation.” But as a college kid I was influenced by Linus Pauling, who said, “We came from chemistry.” I have spent my career trying to get a chemical explanation for life, the explanation of why we are human beings and not monkeys. The reason, of course, is our DNA. If you can study life from the level of DNA, you have a real explanation for its processes. So of course I think that the human genome project is a glorious goal.

People ask why I want to get the human genome. Some suggest that the reason is that it would be a wonderful end to my career—start out with the double helix and end up with the human genome. That *is* a good story. It seems almost a miracle to me that fifty years ago we could have been so ignorant of the nature of the genetic material and now can imagine that we will have the complete genetic blueprint of man. Just getting the complete description of a bacterium—say, the five million bases of *E. coli*—would make an extraordinary moment in history. There is a greater degree of urgency among older scientists than among younger ones to do the human genome now. The younger scientists can work on their grants until they are bored and still get the genome before they die. But to me it is crucial that we get the

human genome now rather than twenty years from now, because I might be dead then and I don't want to miss out on learning how life works.

Still, I sometimes find myself moved to wonder, Is it ethical for me to do my job? A kind of backlash against the human genome project has cropped up from some scientists—good ones as well as not so good ones. What seems to have outraged many people was that, in 1990, against the proposed increase of 3.6 percent in the president's budget for all NIH funds, the human genome project was proposed for an increase of 86 percent—from roughly \$60 million to \$108 million. Feeling dispossessed, some scientific groups have begun to behave like postal workers' unions. The biological chemists, the molecular biologists, and the cell biologists have hired a lobbyist, a former congressman from Maine, to get the overall NIH appropriation increased. If such moves succeed, then maybe we won't have this terrible situation of really good scientists claiming that they are not getting funded because all the money is going to the human genome project.

In the meantime, hate letters have made the rounds, including the rounds of Congress, contending that the project is “bad science”—not only bad, but sort of wicked. The letters say that the project is wasting money at a time when resources for research are greatly threatened: If good people are failing to get grants, why go ahead with a program that is just going to spend billions of dollars sequencing junk? In 1990, someone in my office tried to get a distinguished biologist to help peer-review a big grant application. The biologist said, “No, not the human genome!” as though he were talking about syphilis.

The biologist sent me a FAX asking me to explain why he should not oppose the human genome program. I called him up and said that, though I couldn't prove it, Congress actually seemed to *like* the human genome program because it promised to find out something about disease. Congress was excited that maybe we scientists were worried about disease instead of just about getting grants. The primary mission of the National Institutes of Health is to improve American health, to give us healthier lives, not to give jobs to scientists. I think that the scientific community, if it wants to be ethically responsible to society, has to ask whether we are spending research money in a way that offers the best go at diseases.

The fact is that understanding how DNA operates provides an enormous advantage over working only with proteins or fats or carbohydrates. The best illustration of this advantage has been tumor viruses. If we had not been able to study cancer at the level of the change in DNA that starts it, the disease would still be a hopeless field. Every time a new enzyme was discovered, hope would rise that it was the cause of cancer. Cancer used to be considered a graveyard for biochemists, even good ones, many of whom wanted to cap their careers by solving cancer but failed. Not until the genetic foundation for cancer was identified could you really begin to say what goes wrong to make this terrible human affliction.

A similar example is Alzheimer's disease. Are we going to find out what Alzheimer's is and why it causes brain failure without getting the genes that we know predispose certain people to the disease? Maybe we will, but I would not bet on it. But if we can get the gene or genes implicated in the disease, I am confident that we will save hundreds of millions of dollars, if not billions, that would have been spent on worthless research.

Every year, Congress passes a bill for even more money to study Alzheimer's. Congress is voting for good goals, but we do not really know how to use the money. It is not as if all the federal budget for health and all the basic research grants add up to good research. All the study sections in the National Institutes of Health do not receive applications of equal value; they often endorse research projects or programs because they address important problems. The programs themselves are not terrible, but they often have a low probability of paying off. I am sure that half the NIH budget is spent on good intentions rather than on a realistically high probability that a research program will have a direct impact on one of the major human diseases.

The pressure is enormous to do something about mental disease because it can be terrible, as anyone knows who has a friend or family member suffering from it. We do spend a vast amount of money studying mental diseases, yet the effort yields very little. Manic-depressive disease leads to great moments of mania—perhaps the successful careers of a number of scientists can be attributed to it—but it also leads to depression, tragedy, and suicides. Lithium relieves some of the symptoms, but a drug is not the complete answer, as any psychiatrist will tell you. It is pretty

clear that manic depression has a genetic cause. Several scientists thought they had located the gene on a chromosome. But then it got lost, and so long as it is lost, we are lost.

It is also pretty clear that alcoholism bears some relationship to genes. This view comes from studies on identical twins adopted and raised by different families. There *are* alcoholic families. It is not likely that their members are morally weak; they just cannot tolerate alcohol chemically. But no one has found the gene or genes for susceptibility to alcoholism, and the chance of finding the genetic sources are probably low until a much more sophisticated human genetic community exists—plus the money to get the pedigrees and all the genetic markers.

Some diseases are not going to be easy to crack. For a long time, people have been trying to discover the cause of schizophrenia by looking for chemical differences in the urine or the blood, a research strategy that has not been successful. It is not going to be easy to find the genes behind schizophrenia either, because reliable pedigree data are difficult to compile and the condition is hard to diagnose. Thus both directions offer low probabilities, but it is still better to waste your money doing genetics because genetics lies at the heart of so much. Of course scientists should find out what the brain is. I believe in neurobiology and have tried to help raise money to support the field. But I do not believe that its current approaches will necessarily lead to the real, deep cause of manic-depressive disease.

In 1989 Congressman Joe Early said to me, "I'm tired of putting fingers in dikes!" In combating disease, genetics helps enormously if it is a bad gene that contributes to the cause. Ignoring genes is like trying to solve a murder without finding the murderer. All we have are victims. With time, if we find the genes for Alzheimer's disease and for manic depression, then less money will be wasted on research that goes nowhere. Congressmen can only feel good if they are spending money on good things, so we have to convince them that the best use for their money is DNA research.

The human genome project is really trying to push a little more money toward DNA-based research. Since we can now produce good genetic maps that allow us to locate culprit chromosomes and then actually find the genes for disease (as Francis Collins found the gene for cystic fibrosis), genetics should be a very high

priority on the agenda of NIH research. We are extremely lucky that when James Wyngaarden was director of NIH, he saw to the establishment of what is now a permanent division within NIH called the Center for Human Genome Research. I doubt that I convinced the biologist who sent me the FAX, but I may eventually, since he is very bright. I want to convince as many people as I can of the merits of the human genome project, but not to cap my career and have something that sounds good in my obituary. I can make best use of my time by trying to mobilize the country to do something about diseases that have hit my family and many others. I am sort of a concerned parent for whom things have not gone completely right. So, I am trying to enlist a group of people who will help us get these genes, and do what I think Congress wants us to do.

The ultimate objective of the human genome program is to learn the nucleotide sequence of human DNA. We want the program completed in roughly fifteen years. By completed we do not mean every last nucleotide sequence. If we get 98 percent of the regions that are functional, that will probably be the end of it. We will not worry about spending infinite amounts of money trying to sequence things we know probably contain little information. We could define the end of it to be the identification of all the human genes—that is, we will be done when we have located the coding sequences and can declare that human beings on the average contain, say, 248,000 genes, with variations such that some individuals, for example, have a gene present in four copies and some in three, and that for some the gene is nonessential. It has recently been learned that only a third of yeast genes are essential. Knock out two-thirds of them and the yeast still multiply. Studying things that are not essential will keep the people in the yeast world going for a long time. I think we can safely say the project will be over when we can identify the genes.

We probably will be unable to identify the genes until we get most of the DNA sequenced, because we will not know where they are. It would be nice if the whole program could be done by copy DNA (cDNA)—that is, by purely functional DNA—so that we would not have to sequence all the junk, but we will never know whether we have all the cDNAs. This is not to say we

should not do cDNA; we will actually fund grants for people trying to find better techniques for getting rare cDNA in tissue-specific places. But I think that we have got to sequence the whole thing.

In the first five years, we will push to achieve three major objectives. First, we will try to get good genetic maps, so that each chromosome has enough genetic markers on it actually to locate a gene if a pedigree is available. Currently, we have only about 150 markers that are sufficiently informative for assigning the location of genes. We have started a crash program to persuade people to make a lot of markers and to put them into a public repository made available to the whole world. We want to change the current practice among researchers of not sharing their markers because they want to be the first to find a gene and encourage everyone to make markers available to everyone.

The second objective is to make overlapping fragments of DNA available so that anyone looking for a gene in a particular piece of a certain chromosome will be able to get it by paying some nominal sum. The fragment will not be totally free, but it will certainly be there for anyone who seriously wants it. Techniques for doing this seem to be available now; it should not require more than \$10 million to stockpile overlapping fragments of a given chromosome. To put this figure into perspective, Francis Collins has said that finding the cystic fibrosis gene was expensive—between \$10 million and \$50 million. If all the markers had been available, it would have cost only \$5 million. I think we can establish an overlapping fragment library for the entire human genome for a couple of hundred million dollars, which will certainly reduce the costs of subsequent disease hunts. We will end up with a map of overlapping fragments, each one identified by three or four DNA sequences along it called sequence tag sites. With PCR, researchers will be able to pull out all the human DNA that may be wanted.

The third major objective is to support scientists trying to do megabase sequencing in one place in a reasonable period of time. An example of this type of project is a proposal from Walter Gilbert to sequence a mycoplasma, which is really a small (800 kilobases) bacterium. Gilbert's proposal, whether he lives up to it or not, is to do a million bases a year within two years. We want to encourage people to do sequencing of megabases with the aim of

reducing the cost—so that within a couple of years it will fall to about a dollar a base pair, and then perhaps even to fifty cents. We will not accept a grant application from someone who proposes to sequence some DNA the old fashioned way, with graduate students or postdoctoral fellows, at the current cost—five to ten dollars a base pair—just out of curiosity about it.

People continue to work in the old-fashioned way, but I have my doubts that it advances careers. It used to be that you could get a job if you could sequence DNA. Now, if you sequence too much, you probably cannot get a job because you have done nothing interesting. We human genome projecters are actually *good* people; we want to save graduate students and postdocs from ever having to sequence by giving them a tool. We want sequencing to be done by much cleverer ways—by machine or by multiplexing and with automatic gel readers—so that researchers will not have to go crazy just doing the same sequencing procedures over and over.

A Japanese scientist told me a very unlikely story—one so unlikely that it must be true. He was describing the Japanese effort to sequence a chloroplast DNA, which was about 120,000 base pairs. Two groups were in a competitive race in Japan to get the sequences of a few different chloroplasts. Both came out successful, but mutiny broke out in one of the teams. It is imaginable that an American graduate student might tell his supervisor to go to hell; it is unimaginable that a Japanese graduate student might do the same. In the face of the extraordinary mutiny, the Japanese supervisors decided that forced-labor sequencing was too inhumane and resolved to change the system.

We hope to spend 10 to 20 percent of our total money trying to develop sequencing methods that could make the life of future students more humane. We face the problem of convincing NIH study sections—those peer-review bodies that assess and approve research proposals—to take a sufficiently adventurous attitude toward the development of fast-sequencing technologies. They tend to be willing to fund something only if they know it can be done. What we have to fund are projects whose outcome is uncertain, and we know no way to proceed other than to trust the investigator who has a good idea and to give out the money. Since we have not yet done a megabase in a single project, that makes for a problem in obtaining study-section approval. In con-

trast, mapping will breeze through peer review because many scientists have shown it can be done. I am confident that with all the brains in our field, we can reduce the cost of sequencing by a factor of ten.

The NIH genome project will also try to get some real data on model organisms. I will be happy if we get ten quite different bacteria sequenced up through yeast. We are now supporting a joint program between the Medical Research Council, in England, and the Laboratory of Molecular Biology in Cambridge, and the group in St. Louis that has developed yeast artificial chromosomes to sequence the genome of a roundworm. The roundworm community is eager to do it because they've already got the overlapping DNA fragments. We hope to get the sequence out in ten years. It's about the equivalent of an average human chromosome—about a hundred megabases—but with less repetitive DNA, and so probably with fewer problems. There is also an effort to sequence a plant genome, *arabidopsis*, which we hope will be led by the National Science Foundation with help from other agencies, including ourselves. This is roughly seventy megabases, and the project should be a real boon to botany. Except for perhaps one bacterium, none of this probably would ever have been funded in the absence of the human genome program.

Among the reasons for wanting to find bacterial genes is to help find the human ones. People ask, How are you going to identify a gene if it is interspersed with so much junk and you lack a cDNA? How are you going to know you have it? That is obviously going to be hard in some cases, but if you have obtained the corresponding bacterial gene without many repetitive sequences and if you are clever, you ought to be able to spot the differences. I can imagine that typical work for undergraduates will be to find the gene once all the sequence has been obtained. Professors could tell their students: If you can identify a gene, we will let you go on to graduate school and do real science.

The human genome project is sufficiently justifiable so that if no other country wants to help fund it, the United States should do the whole thing. We are rich enough to do it. But I doubt that we will be allowed to do it alone, because others are going to worry that it might actually be commercially interesting, and they will worry that we will be disinclined to distribute the data very fast if we have paid for it ourselves. It is my hope that we can

spread out the cost of sequencing and data distribution over many countries. As soon as a gene has been identified, it should be thrown into an international data base.

But there are problems that I don't see how to get around. If a stretch of DNA is sequenced in an academic laboratory, a university lawyer will say, "That looks like a serotonin receptor. Patent it!" Mutant forms of the cystic fibrosis gene have been patented by the universities of Toronto and Michigan. They will get some royalties and maybe build better student unions with the revenues. I am at a loss to know how to put valuable DNA sequences in the public domain fast when a lot of people want to keep them private. I just hope that other major nations come in. The Japanese will not let anyone who doesn't pay for it see their work. I figure that strategy might work. People might actually pay for sequence information if that is the only way to get to see it. So I have to seem a bad guy and say: I *will* withhold information that we generate if other countries refuse to join in an open sharing arrangement. But, in truth, it would be very distasteful to me to get into a situation where we were withholding the data for reasons of national advantage.

The acquisition of human DNA information has already begun to pose serious ethical problems. I think that somehow we have to get it into the laws that anyone's DNA—the message it gives—is confidential and that the only one who has a right to look at it is the person herself or himself. Still, the ethics get complicated if you can spot a gene in a newborn child that produces a disease for which no treatment exists. Sometimes these defects will be hard to spot, but sometimes, as in muscular dystrophy, they can be very easy to detect. As we begin to get data of this kind, people are going to get nervous and some are going to be violent opponents of the project unless they can feel that they or their friends will not be discriminated against on the basis of their DNA. If someone can go look at your DNA and see that you have a deletion on one of your anti-oncogenes and that you will be more liable to die of cancer at an early age, then you might be discriminated against in, say, employment or insurance coverage.

Laws are needed to prevent genetic discrimination and to protect rights that should not be signed away too easily. If you are poor, it will be highly tempting to say, "Yes, look at my DNA because I want the job in the asbestos factory." If you have no

money, a job in an asbestos factory is better than no job. Issues like these demand a lot of discussion, at least so that DNA-related laws are not enacted prematurely. For that reason, we are putting more than 3 percent of the genome project money into an ethics program; and we will put more into it if we find that it needs more.

We have faced up to this challenge already with DNA fingerprints. The National Center for Genome Research has given \$50,000 to the National Research Council–National Academy of Sciences study on DNA fingerprinting, which has lawyers and judges advising it. The police want a DNA register of sex offenders; other people may want one of dishonest accountants. People will want DNA fingerprints to prove that a politician's children are really his. At a meeting in Leicester, England, Alec Jeffries showed a slide of a letter from a woman who runs a small hotel in Wales and who wrote that it would be a good idea to have a DNA fingerprint register of bedwetters. Different people will want different information—the possibilities are unlimited. I don't think *anyone* should have access to anyone else's DNA fingerprints.

We need to explore the social implications of human genome research and figure out some protection for people's privacy so that these fears do not sabotage the entire project. Deep down, I think that the only thing that could stop our program is fear; if people are afraid of the information we will find, they will keep us from finding it. We have to convince our fellow citizens somehow that there will be more advantages to knowing the human genome than to not knowing it.