DNA. The metazoan with the smallest known genome is the worm *Caenorhabditis*, which has 8×10^7 base pairs; another small genome is that of the fly *Drosophila*, with 1.7×10^8 base pairs. These examples show that a complex organism *can* be constructed with only 3 to 6% of the DNA found in humans. Interestingly, even in these organisms a considerable fraction of the DNA does not code for proteins.

At this point one might be tempted to say flies and worms are one thing, but obviously humans have more DNA because of their greater complexity. In fact, among eukaryotic organisms genomic DNA content has virtually nothing to do with complexity. Here we must deal briefly with what is known in the chromosome field as the C-value paradox (C-value is another name for genome size). The paradox has two parts (*1*). The first is that organisms of similar morphological complexity for evolutionary relatedness often have vastly different C-values; the second is that most eukaryotes have much higher C-values than can be accounted for by protein-coding needs, humans being merely one of many examples. Among vertebrates the highest C-values belong to some salamanders, which have about 30 times as much DNA as humans; surely salamanders are not 30 times more complex than humans! On the other hand, some fish manage with a genome less than a third that of humans. Among invertebrates one finds the same puzzling variation. As already noted, *Drosophila* has a small genome, but this is not because it is an insect; some grasshoppers have two to three times as much DNA as humans. Plants have a similar range of values, again not related to evolutionary or morphological criteria (lettuce has much less DNA than humans, but corn and lilies much more). In summary, therefore, the human genome is large relative to its protein-coding needs, but as genomes go, it is neither very small nor very large.

Over the past 20 years the question of genome organization, including the C-value paradox, has attracted enormous attention, both experimentally and theoretically. Perhaps the most important generalization is that variations in genome size are *not* due to variations in the reiteration frequency of protein-coding genes. Thus, the idea that organisms with high C-values have many copies of each gene, whereas those with low C-values have only one or a few, is certainly false. There *are* a great many reiterated sequences in organisms with high genome contents, but few of these code for protein.

Where is the noncoding DNA? Most of it is in "spacer" regions between genes, although a minor and variable amount is within genes as introns. Neither the spacers nor introns (with a few exceptions) code for proteins, and there is no evidence that their specific sequences are important, as opposed to their length, position, secondary structure, or some other feature. If one wants to argue that we should sequence 90 to 98% of the human genome in hopes of discovering some new sequence-dependent function of introns and spacers, the answer is simply that that is bad science. There are already plenty of such sequences stored in computers, and if one wanted another million or so bases for analysis, they could be had cheaply without sequencing the whole genome. Until the spacer and intron DNA's are shown to have some sequence-dependent role, there is no intellectual justification for sequencing them at random.

Although I strongly oppose the sequencing project in its simplistic version, I do believe that knowledge about the human genome is intrinsically interesting and certain to be of medical value; furthermore, we have the techniques and an adequate theoretical framework to justify greater effort in this area. I believe we should proceed simultaneously along two lines. First, mapping studies could begin, using as a guide what Alan Coulson and John Sulston have already accomplished with the worm *Caenorhabditis* (*2*). Even this task will be heroic, since the human genome is 40 times larger than the worm's (and mapping requires all of the genome, coding and noncoding). Preliminary chromosome sorting would reduce the problem a great deal. Second, individual investigators should continue to sequence whatever genes appear to be of greatest interest. If a larger scale project is undertaken, then it should begin with complementary DNA (cDNA) clones. In these clones most of the DNA codes for protein and therefore is currently interpretable. Furthermore, the cDNA clones could be matched to their appropriate places on the physical map by nucleic acid hybridization. If it seemed valuable one could then sequence the genomic regions corresponding to the cDNA's. An enormous advantage of this approach is that one would already know the limits of each gene as well as the correct reading frame, information that is difficult to extract from raw and inevitably inaccurate sequence data in an uncharted region of the genome.

The mapping and cDNA sequencing would be expensive. After the initial strategy was worked out in detail, the intellectual challenges might not seem so alluring. Thus the work might well require some kind of contractual or programmatic aspect outside the usual investigator-initiated grant system. However it may be organized, my plea is simply that we think about this project in light of what we already know about eukaryotic genomes and not set in motion a scientifically ill-advised Juggernaut.

JOSEPH G. GALL
Department of Embryology, Carnegie Institution of Washington, Baltimore, MD 21210

REFERENCES

1. J. G. Gall, *J. Cell Biol.* **91**, 3s (1981); T. Cavalier-Smith, Ed., *The Evolution of Genome Size* (Wiley, New York, 1985).
2. A. Coulson, J. Sulston, S. Brenner, J. Karn, *Proc. Natl. Acad. Sci. U.S.A.*, in press.

Underground Storage Tanks

One critical point about the recent briefing by Marjorie Sun "EPA grapples with regulating underground storage tanks" (News & Comment, 1 Aug., p. 518) should be clarified.

The Environmental Protection Agency (EPA) national survey on underground petroleum storage tanks, released on 24 June, makes no estimate and draws no conclusion about the amount of the nation's ground water, including drinking water, that may be at risk from tank leaks.

The EPA survey specifically emphasized that while it found 35% of the tanks tested failed a tank tightness test, this does *not* indicate those tanks are leaking under normal operating conditions.

A tightness test is a screening mechanism. It must be followed with corroborating testing procedures to avoid confusing leaks with other factors that could cause a test failure—often loose fittings or worn gaskets at or above the top of an underground tank.

At least three major oil companies that have used the same type of underground tank tightness test as the EPA found, when they completed follow-up testing procedures, that actual leak rates were vastly lower than test failure rates. Leak rates for these three companies ranged from 0.97 to 2.6%, whereas original tightness test failures ranged from 10 to 19%. The EPA, unfortunately, did not undertake any follow-up procedures.

Even when leaks occur, they typically are detected and corrected before ground water is affected and usually are confined to the property of the tank system owner. The data from the member companies of the American Petroleum Institute make it clear that the vast majority of their tanks—95% or more—are not leaking and do not represent a major threat to drinking water.

WILLIAM F. O'KEEFE
American Petroleum Institute, 1220 L Street, NW, Washington, DC 20005