

suitable vegetation of an oceanic impact that could still ignite wildfires. An impact may have provided the energy necessary to bring oxygen from a (possibly metastable) level of 24% back down to 21%, a value that seems to reflect a balance between photosynthetic production of hydrocarbons and combustion of terrestrial vegetation.

MARK A. S. McMENAMIN  
DIANNA SCHULTE McMENAMIN  
*Department of Geology and Geography,  
Mount Holyoke College,  
South Hadley, MA 01075-1484*

#### REFERENCES

1. P. C. De Graciansky *et al.*, *Nature (London)* **308**, 346 (1984).
2. P. Cloud, *Paleobiology* **2**, 351 (1976).
3. A. J. Watson, J. E. Lovelock, L. Margulis, *Biosystems* **10**, 293 (1978).

### Nucleic Acid Database Management

We would like to comment on the database management model presented by the nucleic acid database groups at the joint National Institutes of Health-European Molecular Biology Laboratory workshop "Future databases for molecular biology," recently held in Heidelberg. This model was designed to cope with the explosive increase in the amount of nucleic acid sequence data expected in the coming years. The model should be commended because it calls for the direct participation of the research community in the input and verification of the data. The plan is for the database groups to develop and provide software that allows the investigators to prepare their own data for submission to the databases; quality control mechanisms that automatically check the data will be designed into the software.

Under present policies the preparation and verification of data is the most time-consuming task; because this work is performed by research scientists, it is also expensive. By relegating the role of data preparation to the research community and concentrating on software development, the database groups hope to develop a management system that can cope with the increasing amount of data at minimal cost. For the past 10 years, we have been using software to maintain the quality of the National Biomedical Research Foundation-Protein Identification Resource Protein Sequence Database. Such software can save a considerable amount of time and reduce staff. However, it has limitations, in that it can check only those aspects of the data specified by the designer.

Scientific information is fundamentally different from the data usually treated by

computer science in that its nature changes as the understanding of the research community increases. New data often present new problems, and the software must be continually upgraded. We fear that if the nucleic acid database groups stop examining the data, they will not know what changes need to be implemented in software and the resultant software will not be up to the task.

The participants in the workshop presented a visionary picture of the use of future sequence databases as powerful research tools; for example, they could be used to identify relations between sequences, predict structure and function, and allow the detailed chemical mechanisms of gene control and activation to be modeled. Powerful organizational structures could map the interrelations between the data and allow these mappings to be used in generating new information. The resulting information resource would be more than just the sum of its parts. Such uses require highly organized databases in which the relations between each entry and the rest of the database are clearly defined. Quality control in such a structure involves not only validating the information in each entry, but also ensuring the correct representation of the interrelations between the data. The model presented by the nucleic acid database managers does not contain a mechanism for generating or maintaining these organizational structures.

Contributing research scientists most certainly are more knowledgeable about their data than are the database producers; often, however, the scientists do not appreciate the relation of their data to the overall structure of the database. That is the role of those maintaining the database, whose task will consume an increasing amount of our resources.

It was brought out quite clearly at the workshop that the amount of money being spent to maintain the databases is only a tiny fraction of that being spent to generate the data. We must seriously ask ourselves why we are willing to spend so much money generating data but are unwilling to spend enough money to ensure that the data are properly stored and organized and that all possible information is extracted. If the databases can become the powerful research tool envisioned by the workshop participants, it will certainly be worth the investment.

DAVID G. GEORGE  
LOIS T. HUNT  
WINONA C. BARKER  
*Protein Identification Resource,  
National Biomedical Research Foundation,  
Georgetown University Medical Center,  
Washington, DC 20007*

### Saturated Fat Avoidance

I would like to clarify the quotes of what I said to Gina Kolata on the subject of the avoidance of dietary fat (Research News, 23 Jan., p. 436). In countries such as the United Kingdom or the United States, replacement of much of the saturated fat from dairy and other land animal products by other commonly available foodstuffs will reduce plasma low-density lipoprotein cholesterol, which in turn will reduce the likelihood of premature death from coronary heart disease. Such changes are of substantial value and are incorporated into many sets of dietary guidelines. Some experts believe such changes will also help avoid cancer, but the evidence in this respect is less secure than the evidence that they will help avoid coronary heart disease. Hence, I said that when such changes are recommended then it should chiefly be because they *will* help avoid heart disease rather than because they *may well* avoid cancer. Whether or not they also have any substantial effect on cancer, the determinants of plasma low-density lipoprotein cholesterol are collectively of an importance comparable with that of tobacco, and collectively they and tobacco still account for about half of all British deaths in middle age.

RICHARD PETO  
*ICRF Cancer Studies Unit,  
Radcliffe Infirmary,  
University of Oxford,  
Oxford OX2 6HE, United Kingdom*

#### Correction

In Bernard D. Davis' article "Bacterial domestication: Underlying assumptions" (Policy Forum, 13 Mar., p. 1329), the word "no" was erroneously omitted from the last sentence of the first full paragraph in column two on page 1334. The sentence should have read, "But this view builds on a parallel between genetic engineering and the physical technologies, rather than on a much closer model, with very different predictions: the domestication of wild organisms, resulting in enormous benefits and no catastrophes."

*Erratum:* The opening sentence of Igor B. Dawid's review (6 Feb., p. 695) of *Gene Activity in Early Development* by Eric H. Davidson was incorrectly printed. The first paragraph of the review should have read as follows. "As in the two earlier editions of Davidson's standard work, much of this new book is devoted to discussion of quantitative work on RNA accumulation and distribution during development. Davidson also discusses in detail spatial regulation of gene expression in the embryo and its relationship to lineage determination, providing a strongly comparative point of view and generating broad and well-balanced interpretations of large bodies of connected facts. The result is an excellent book that represents the field from a personal yet broadly convincing vantage point."