Vignettes: Funding and Non-Funding

Bodies such as the UK research councils, who now pay most of the direct costs of academic science, are becoming more aware of the institutional impact of their award or non-award of grants and contracts. . . . A peer review panel specializing in, say, subconducting monomerization has no right to complain that there are no alpha-rated projects being proposed in this new field, when, as it happens, their own adverse decisions of five years before had caused the only good British researcher interested in developing this subject to pack up and go to Australia.

—*John Ziman*, in *Prometheus Bound: Science in a Dynamic Steady State* (Cambridge University Press)

A scientific puzzle must reach a stage of "solution readiness" before much good can come from an all-out effort to solve it. The puzzles of the age of the earth and the history and future of the sun were not "solution ready" during the years 1860–1890, when Kelvin was most active. This was long before the days of huge government grants for research; but we can confidently say that had there been such grants, handing out millions of pounds sterling to physicists in Kelvin's day for attacks on these problems would not have hastened their solutions.

—*Garrett Hardin*, in *Living Within Limits: Ecology, Economics, and Population Taboos* (Oxford University Press)

One way to lessen the risk inherent in undertaking a major [research] project is to make sure that you spend enough money on it. After a research department or funding agency has invested enough in your goals, it has a real stake in your success and becomes very reluctant to admit that your project is not working out. No one ever got ahead in science by saving money.

—*Peter J. Feibelman*, in *A Ph.D Is Not Enough! A Guide to Survival in Science* (Addison-Wesley)

Genome Informatics

Biocomputing. Informatics and Genome Projects. DOUGLAS W. SMITH, Ed. Academic Press, San Diego, CA, 1993. xii, 336 pp., illus. $49.95 or £38.

The Human Genome Project has always known that it needed informatics. The 1987 report of the Health and Environmental Research Advisory Committee that originally defined the project for the U.S. Department of Energy identified informatics as a major area of concern. The 1990 NIH-DOE five-year plan for the project devoted a section to informatics and established a Joint Informatics Task Force to define the key issues. Adequate informatics components are now viewed as essential features of genome research centers. Even so, there is only the vaguest consensus in the community as to what genome informatics amounts to.

Biocomputing: Informatics and Genome Projects presents genome informatics essentially as it is presented in the 1990 five-year plan: as a set of discrete components that support the biological analyses that compose a genome project. Smith's components are automated laboratory notebooks, DNA and RNA sequence analysis, protein-structure analysis, and database activities. Contributions by Smith *et al.* and by Bruce and Fred Cohen discuss the analysis of DNA and protein sequences, respectively. Smith's chapter focuses on adapting sequence-comparison algorithms to supercomputers and is primarily a review of available parallel architectures. This chapter is not for beginners: familiarity with standard sequence-analysis algorithms is assumed. The Cohens' chapter summarizes a broad range of approaches to structure prediction, from local secondary structure to complete tertiary structure. Three papers address data management issues. Clark *et al.* describe the Genome Notebook database, developed to support the chromosome 11 physical mapping effort at the Salk Institute, and discuss some of the related issues raised by physical mapping projects. Yoshida *et al.* outline an approach to developing mapping databases in the Prolog programming language, providing a brief tutorial in Prolog in the process. Mount and Schatz propose the development of community databases to serve as communication and collaboration-support systems and briefly describe the Worm Community System, developed for the *Caenorhabditis* community by Schatz and colleagues, as an example.

The remaining contributions collectively provide a somewhat broader view of the field and link genome informatics to other aspects of computational biology. The chapters by Henikoff on protein-sequence motifs, Konopka on statistical analysis of DNA sequences, and Wills on methods for phylogenetic analysis connect the analysis of data derived from genome projects to broad evolutionary questions. The final chapter, by Galper and Brutlag, introduces the problem of modeling biological processes. These chapters point up the fact that the division between genomics and the rest of biology is artificial at best, and potentially deleterious to both. Organisms and evolution cannot be understood fully without a good grasp of molecular biology, including genomes and genome-level processes; similarly, a genome cannot be comprehended outside of the functional context provided by cells, tissues, organisms, populations, and evolutionary histories.

What is missing from this book is, in an important sense, what is missing from the field of genome informatics as a whole. There is, on the one hand, no big picture, no synoptic presentation of how the analysis of information relates to the genome project, and to the rest of biology. On the other hand, there is only the most cursory mention, in the paper by Clark *et al.*, of how informatics strategy meshes with biological strategy in planning and executing a large-scale project. Genome informatics, like computing in physics, encompasses activities ranging from quite speculative theory to systems engineering. Prioritizing and integrating these activities has proved to be a major challenge for genome laboratories. The extent to which designing an experimental program is itself an informatics activity goes largely unrecognized. A book meant to "provide the groundwork necessary to understand genome-related informatics," as the publisher describes this one, needs to be more than a collection of brief reviews: it needs to provide a perspective that allows a nonexpert reader to understand both the scientific issues and the practical choices involved in developing an informatics program. In most genome laboratories, the first question about informatics is, "What do we need to do?" This book does not answer that question.

Chris Fields
Institute for Genomic Research,
932 Clopper Road,
Gaithersburg, MD 20878, USA