The biologist and the computer

Brunak, Søren

Published in:
Trends in Genetics

Publication date:
2000

Document Version
Publisher's PDF, also known as Version of record

each field to assess the achievements, and help make them intelligible to the general reader. Her introductory chapter 'A Passion for Discovery' and her 'Afterword' describe the changing patterns of discrimination against women across the 20th century.

Most of her subjects faced enormous obstacles, both institutional and personal, some from a presumption that their work was done by a male supervisor or collaborator, some from male competitors. Some suffered racial persecution or prejudice (Meitner, Levi-Montalcini). Some suffered crippling diseases: Marie Curie and Iréne Joliot-Curie from the effects of radiation, Dorothy Hodgkin from arthritis, Gerty Cori from bone marrow disease in her last decade. These women triumphed because they loved their science passionately and were thrilled by their scientific breakthroughs.

Most came from supportive family backgrounds, half of them Jewish; Jocelyn Burnell is a Quaker (but her husband was less than supportive). Hoyle, Ostriker, and Gold (who gave the first explanation of pulsars) said that the Nobel award should have included Jocelyn Bell, whose persistence led to the discovery of pulsars. More than half raised children, mostly with a sympathetic husband. There is a charmingly quirky portrait of Barbara McClintock, who played tenor banjo in a jazz group and learned Yiddish as a freshman at Cornell.

Given the enormous institutional barriers, it is not wholly surprising that most of the American Nobelists were nurtured by just two institutions, Hunter College, NY, and Washington University, St Louis. Carl but not Gerti Cori was elected to the National Academy of Sciences and the Royal Society before they won their Nobel prize jointly; after this Gerti also was elected to the NAS.

McGrayne's book affords an interesting and useful commentary on the factors in the world of science that still limit the high achievement of which so many women are capable. Legal barriers and 'nepotism' rules have given way to less overt but still powerful constraints on women in science. Fewer now work without salary or position, but the small numbers of women in senior positions in science bear testimony to the limitations on their resources.

McGrayne remarks that of over 300 Nobelists since 1901, only ten have been women. This 3% resembles the current proportion of women in the world's science academies (3.9%), although some are making efforts of redress. There are now 3.6% of women among Fellows of the Royal Society, following the election this year of an extraordinary proportion of new women Fellows, 5 out of 42. The NAS does better with 6.2% overall, having elected an unprecedented 9 women out of 60 new members this year.

Nobel Prize Women in Science is highly readable, and good value in softback at less than $2 per Nobelist, $1.33 with the inclusion (and investigation) of the should-have-beens. For good measure the book contains some excellent photographs, and Rita Levi-Montalcini's mouth-watering recipe for zabaglione coffee ice cream.

Joan Mason
jm148@cam.ac.uk
History and Philosophy of Science, Free School Lane, Cambridge, UK CB2 3RH.

The biologist and the computer

Computational Methods in Molecular Biology (New Comprehensive Biochemistry, Vol. 32)
edited by Steven L. Salzberg, David B. Searls and Simon Kasif
Elsevier, 1999. $59.00 pbk (371 pages) ISBN 0 444 50204 1

In July 1966 a *Science Magazine* editorial on the human use of computing machines stated that 'The relation of man and computers has entered a new era, in which interaction is becoming quick and simple'. More than 30 years later, in 1999, few would agree that even the use of ordinary personal computers has become that trivial. And it is certainly not true in the emerging field of bioinformatics, where new computational techniques are rapidly changing the field of biology and all its associated disciplines.

Classification and prediction based on sequence data are inherently difficult because sequences, or sequence segments, belonging to a given functional or structural category might appear very different, while, at the same time, examples belonging to different categories might be very similar. This situation calls for powerful, non-linear complex algorithms that can cope with the vast diversity created during evolution. In a new volume of *Computational Methods in Molecular Biology*, editors Salzberg, Searls and Kasif have brought together an impressive selection of papers, all authored by leading bioinformaticians. The volume covers this large multi-disciplinary field of research quite well, from sequence alignment, pattern discovery, gene finding to protein structure modeling and prediction. One important new area that has been left out is techniques for the analysis of gene expression data originating from DNA-array experiments.

The volume contains a nice tutorial on computation for biologists, in which basic concepts are introduced and jargon is defined. It includes a warning against the usual computer-science propaganda on algorithms being computationally intractable, when approximations can often lead to fast methods that return biologically relevant knowledge. Many bioinformatics applications have been based on algorithms that, in principle, are intractable (presumably non-polynomial time computable), yet a large number of very useful prediction schemes have been created in this manner. Training of a multi-layer neural network is intractable, but, driven by sufficient amounts of experimental data, we have seen that they can crunch a lot of sequences and predict quite effectively: protein subcellular compartments; signal peptide sequences; protein secondary structure; and intron-exon splice junction, to mention a few.

It is essential that new sequence-analysis tools are developed that are able to go beyond the position-by-position comparison of sequences, which is normally done by alignment and weight-matrix techniques. The excellent chapter by Burge on dependencies in pre-mRNA splicing signals is a good example of such work, which might increase our understanding of the specificity of splicing through the study of the compositional properties of known splice-signal sequences. The analysis reveals a positive association between pyrimidine nucleotides at adjacent positions in the poly-pyrimidine tract, which is putatively related to the affinities of pyrimidine-tract-binding proteins. These techniques can be applied equally well to other types of nucleic acid signals, such as those involved in transcription and translation.

A few chapters have misleading titles that are too broadly formulated, where one is expecting a review of a given area, and not just a description of a single computer program. In the contribution from Xu and Uberbacher on gene prediction using neural networks, the reference list contains only a handful of citations to this type of work, which is misleading given the large literature on

Soren Brunak
brunak@cbs.dtu.dk
Center for Biological Sequence Analysis, Dept of Biotechnology, The Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark.

the subject. However, this is an exception, as most of the chapters present excellent reviews of the subject they cover.

Without the computers (and bioinformaticians to control them) it would make little sense to sequence hundreds of genomes and to produce massive amounts of gene expression data in DNA-array experiments. The data explosion has turned the relationship between experimentalists and bioinformaticians upside down. In the past, it was often necessary to exert considerable amounts of persuasive powers if collaboration with bench researchers was to be established. Now, the problem is one of selecting between projects, and finding areas where computational analysis can optimally complement experimental work. At this point, the relation between the biologist and the computer is far from simple – and due to the complexity of the data it might never be – but a book like this is certainly a competent step in the right direction.

Master designs

Of Flies, Mice and Men
by François Jacob. Translated by Giselle Weiss
Harvard University Press, 1998. £14.95 hdbk (158 pages) ISBN 0 674 63111 0

Max F. Perutz
jhcl@
mrc-lmb.cam.ac.uk

MRC Laboratory of
Molecular Biology,
Cambridge,
UK CB2 2QH.

François Jacob made his scientific name by discovering how genes are switched on and off in bacteria and became known as a writer for his moving autobiography *The Statue Within*. In his latest book, *Of Flies, Mice and Men*, he asks what decides whether a fertilized egg develops into a mouse, a fly or a human. 'What is ... wonderful about the appearance of a new human being is not the nature of the receptacle in which the first stage takes place. It would not even be the accomplishment of making the entire development take place in a test tube. The incredible thing is the process itself. It is that the meeting of the sperm with the egg initiates a gigantic set of chemical reactions, hundreds of thousands of which follow each other, overlap and cross each other in an orderly network of unbelievable complexity. All this to result ... in the appearance of a human baby and never a little duck, a little giraffe or a little butterfly.'

How can we discover what decides the differences between the development of different species? The first clue came from the humble fruit fly. There are flies that grow legs on their heads in the place where they should have antennae. The mutant gene responsible for this monstrosity belongs to a family of genes that determines the fly's body plan. Are these genes unique to the fly? 'There was hardly a chance of finding these genes in organisms other than insects, seeing how different their embryonic developments are. But people looked for them all the same – just to see – and they were stunned. They found them. Everywhere. First in a frog, then a mouse, then in man, in a leech, in a worm … In short, one finds a group of genes very similar to those of the fly in all animals. Everywhere, their role seems to be the same: to define the identity of different cells along the axis from the front to the rear of the animal. If one takes a mutant fly which lacks one of these genes and inserts in its place the homologous gene from a mouse, it works, and it fulfils the same function as the normal fly gene.'

If the development of eyes is initiated by the same gene in humans and flies, then why are they so different? Walter Gehring suggests that in human beings the single *eyeless* gene switches on a cascade of as many as 2 500 other genes. They would code for 2 500 different proteins, whose complex interplay would then govern the growth of the eye. Some of these proteins might be common to human beings and flies and others different. We know as yet next to nothing about them. Will we ever be able to unravel these genes' labyrinthine workings?

Jacob wonders if there might indeed be a limit to the degree of complexity that we can comprehend, such as the interactions between thousands of genes or between billions of neurons in our brains. Jacob fears that 'the human brain may be incapable of understanding the human brain'. I share his fears. He holds that much of evolution has arisen from Nature's tinkering. Nature makes new genes that code for proteins with new functions by putting together bits and pieces from existing genes in new ways, or simply by replacing bits and pieces in existing genes. 'The whole of the living world looks like some kind of giant erector set. Pieces can be taken apart and put together again in different ways, to produce different forms. But fundamentally the same pieces are always retained.' Jacob's book is masterly in combining erudition, wit and wisdom. It clearly describes what we know about the laws that determine animal development – and what we do not know.