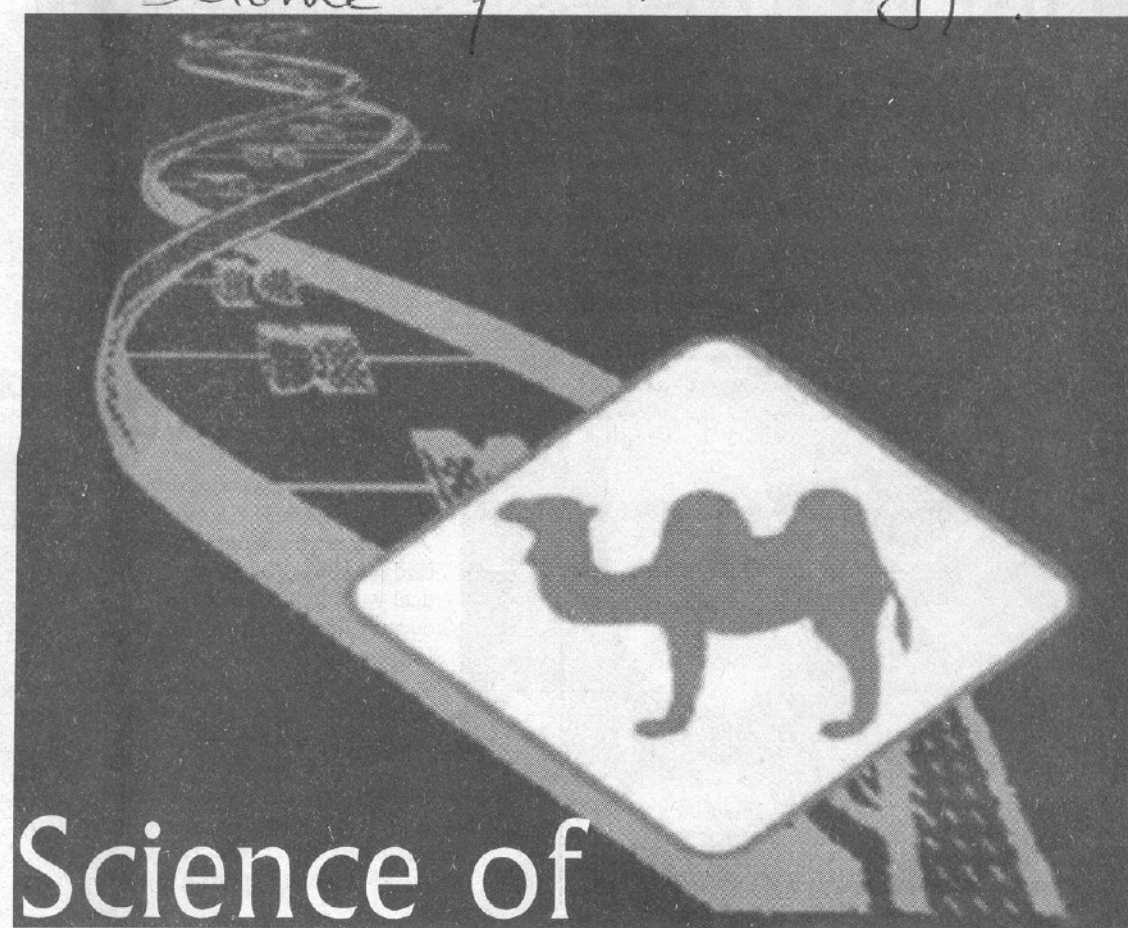


**O**ur knowledge of biology has exploded in such a way that we need powerful tools to organize the knowledge itself. The questions we are asking of biological systems and processes today are getting more sophisticated and complex so that we cannot hope to find answers within the confines of unaided human minds alone. Bioinformatics is an emerging scientific discipline representing the combined power of biology, mathematics, and computers. *Bioinformatics is needed to handle the enormous amount of data being generated by researchers identifying the lengthy DNA sequences of humans, plants, animals, and microorganisms — life's blueprint — and other biological data.* Stored digitally, in computers worldwide, are trillions of pieces of information generated by emerging technologies in molecular biology. The amount of data generated in the form of DNA sequences, by public and private sequencing projects, doubles every 12-14 months and will increase even more dramatically in the coming years. The resulting bottleneck — a wedge between having data and knowing what the data mean — must be overcome to facilitate breakthroughs in medicine, agriculture, and environmental sciences.

One of the most apt definitions of bioinformatics is what has been given by a Professor at Yale University which says "bioinformatics is conceptualizing biology in terms of macromolecules (in the sense of physical chemistry) and then applying informatics techniques (derived from disciplines such as applied maths, computer science, and the statistics) to understand and organize the information associated with these molecules, on a large scale".

Bioinformatics can generally be defined as the study of how information technologies are used to solve problems in biology. The precise definition of bioinformatics is a matter of some debate. The most narrow usage of the term refers to the creation and management of biological databases in support of genomic sequences. The most broad usage includes essentially all applications of computers and information sciences to problems in biology. Of course, bioinformatics is based on the fundamental paradigm of molecular biology: genetic information is stored in sequences of DNA bases (a four letter alphabet), which get translated into sequences of protein



## Science of the future

**DR HABIB BOKHARI** writes about latest developments in bioinformatics

amino acid building blocks (a twenty letter alphabet). Protein sequences have a remarkable ability to reproducibly fold into a three-dimensional shape, and this shape confers on them the ability to form a variety of critical functions for life: enzymatic catalysis, structural support, generation of motion, reception of signals between cells, and transduction of forces (light, pressure, shear) into chemical signals, to name a few.

The functions of proteins combine within a cell to create a living apparatus. Unicellular organisms (such as bacteria) and multicellular organisms (such as people) then function in their environment to acquire nourishment and reproduce. Over time, the demands of the environment create a pressure for these organisms (or their offspring) to build proteins that are better suited to compete for resources. This leads, in turn, to changes in the DNA code in response to environmental pressures, and through Darwinian

natural selection. Perhaps the best way to define the field is to outline the scope of topics covered in most of the principle scientific meetings in the field so far (the International Conference on Intelligent Systems for Molecular Biology<sup>1</sup>, the Pacific Symposium on Biocomputing<sup>2</sup>, and the Annual Conference on Computational Biology<sup>3</sup>).

Biology is still dominated by experimentalists who do not always acknowledge the importance of computation as a research area within biology. "If you don't do bench experiments, you are not a real biologist" is a frequent refrain. Some biologists consider bioinformatics investigators to be important primarily as providers of biologically relevant computing services, and do not acknowledge the set of core paradigms that guide bioinformatics research. The independent status of bioinformatics as a discipline will stand on the success of the field in defining a clear, separable research agenda, creating administrative

units in academia and industry that support this agenda, and gaining attention and support from public and private funding agencies in the country. The focus of this new discipline is to move from the "wet lab" with its gels, sequences and PCR machines to the "dry lab" of hard ware, soft ware and algorithms. A researcher equipped with an internet-connected computer, a working knowledge of the sequence analysis techniques can provide a surprising insight into the macromolecular architecture of a completed genome. Deriving meaningful knowledge from the DNA sequence will define research through the coming decades to broaden our understanding of biological systems. This enormous task will require the expertise and creativity of tens of thousands of scientists from varied disciplines in both the public and private sectors worldwide in bioinformatics.

Computation and Biology share a common obsession with strings of letters, which are translated into complex 3D and 4D structures. Information and Education is becoming increasingly easy to spread (and hard to control). We are on the verge of begin able to collect data on almost any system at costs of terabytes per dollar. We need to not only teach our students



how to cope with this, but start thinking about how to teach these "intelligent" systems as if they were students. As integrated circuits reach their limits so the next generation of computers may be based on quantum computing and or Biologically inspired. We need to be able to teach our students about this revolution and via the Internet teach any one else listening.

The different cultures of the biological sciences and the computer sciences must be bridged to establish a bioinformatics identity and to ensure that biological scientists, computational scientists and mathematicians can communicate with each other. Thus, the need to train scientists whose primary professional identification and disciplinary affiliation is in bioinformatics must be addressed effectively at all levels from college undergraduate to graduate students. Developing countries most notably India, China, Brazil, South Africa and Cuba are already acquiring capacity in bioinformatics.

Some research teams in developed world are applying their expertise to global health, but it is also desirable that the developing world acquires capacity in bioinformatics to achieve a greater level of autonomy and to permit the application of their unique understanding of developing world problems to the search for solutions. Our policy makers realizing the fact that without having to invest heavily in the technologies used to produce them, developing countries like ourselves can take advantage of these data and apply the power of bioinformatics to local health and agricultural problems.

This urgently requires that they should act decisively and promptly by rendering their support and encouragement to bioinformatics. They should not only think of revising their curriculum at college and university level by incorporating this emerging science as an independent course but also encourage our pharmaceutical industries to support the research-based infrastructure to address our indigenous health-related problems. This will open the new avenues of research-oriented jobs both in academia and industry and will also act as a catalyst to strengthen our withering computer industry. ●

*The writer is Assistant Prof. (Biosciences), at COMSAT Institute of Information Technology (CIIT), Islamabad*

units in academia and industry that support this agenda, and gaining attention and support from public and private funding agencies in the country. The focus of this new discipline is to move from the "wet lab" with its gels, sequences and PCR machines to the "dry lab" of hard ware, soft ware and algorithms. A researcher equipped with an internet-connected computer, a working knowledge of the sequence analysis techniques can provide a surprising insight into the macromolecular architecture of a completed genome. Deriving meaningful knowledge from the DNA sequence will define research through the coming decades to broaden our understanding of biological systems. This enormous task will require the expertise and creativity of tens of thousands of scientists from varied disciplines in both the public and private sectors worldwide in bioinformatics.

Computation and Biology share a common obsession with strings of letters, which are translated into complex 3D and 4D structures. Information and Education is becoming increasingly easy to spread (and hard to control). We are on the verge of begin able to collect data on almost any system at costs of terabytes per dollar. We need to not only teach our students