

A new curriculum

A recent editorial in the journal *Bioinformatics* (Altman, R. B., 1998, 14, 549–550) suggests the design and implementation of a graduate course in bioinformatics. The items include routine programs used by molecular biologists and concepts from computer science. It has been recognized in recent years that there is a great demand for researchers with capabilities in these areas.

Even though the term bioinformatics has been coined recently, scientists have been familiar with many of the course items for several decades now. Today, most of the personnel employed in bioinformatics centers (both in India and abroad) are busy creating, annotating, updating, retrieval and analysis of information in the numerous databases. There is a continual demand in this area because, managing databases that grow exponentially is always challenging for full time computer personnel and novel ideas should be explored with great vigour.

I would like to propose that the course be designed to train students with biology, molecular biology, computer science, statistics and applied mathematics right from the graduate level. This however differs from the editorial in that the emphasis is placed on biology, yet retaining the primary aim at a completely different plane of research. It is a common experience that obtaining expertise in several diverse fields is *rare*. It is difficult to find students of present day computer science take a deep interest in biology. The formidable problem to be overcome in attracting students from computer science – basic and -applied – to biology is the competing lucrative software market.

Recently, it is becoming increasingly difficult for biologists (including molecular and structural biologists) to find jobs; the field may soon lose the potential to attract students who may have exceptional interests in diverse fields and a core interest in biology. The problem

is mainly due to the overwhelming demand for skills in *ab initio* analysis. It has become relatively easy to find scientists with excellent skills in molecular biology, structural biology, and immunology. But it is *rare* to find personnel who can correctly predict *ab initio*, the properties of a molecule from the 1-dimensional sequence. This requires remarkable insight powered by a clear understanding of *all* the areas mentioned above and mathematics.

At present, we find mainly, post-doctoral researchers take exceptional interest in diverse fields. If an intensive training has to be provided in bioinformatics, it should necessarily be at the B Sc/M Sc level. A full 3 years M Sc course could be an option. Alternatively, portions of bioinformatics topics – depending upon the type of ‘core strengths’ available in an institute – can be included as part of the PhD course work programs. Some areas that must be included are—

Pairwise sequence alignments (dynamic programming, heuristic methods, similarity methods); Multiple sequence alignment; Hidden Markov models; Phylogenetic trees; Fragment and map assembly and combinatorial approaches to sequencing; RNA secondary structure prediction; Sequence frame extraction/annotation; Protein homology modelling; Protein threading, protein molecular dynamics; Protein *ab initio* structure prediction; Integration of molecular databases; Support of laboratory biology; Design and implementation of biological database/knowledge bases. Several items from computer science have also been suggested: Optimization; Dynamic programming; Bounded search algorithms; Cluster analysis; Classification; Neural networks; Genetic algorithms; Bayesian inference; Stochastic context free grammars.

The bioinformatics course, however, should not just create a set of skilled technicians. Neither, should it be limited

to training biologists in the use of a few computer programs. Instead, the students should be expected to develop: a deep interest in biology and biological intuition, the ability to visualize sequence and structure using various mathematical approaches, excellent skills in writing computer programs, the ability to explain in ‘simple terms’ the findings and their significance and an interest to use these to address the health and food problems of the nation.

The main purpose of bioinformatics – to my mind – is to solve the numerous complex problems in biology. Experimental confirmation of *in silico* predictions can be done through active research with already experienced investigators who were previously post-doctoral researchers with wide capabilities. This is a growing area, and given the fact that there are very many complex problems in biology, the field would be expected to attract the most interested students.

It has been the experience of very many working in computational biology, that, it is not hard to find exceptions to any of the models synthesized to explain sequence and/or structure evolution. Most often, we require a mathematical approach to address an issue. Because India is reputed for many great mathematicians, molecular and structural biologists, and computer scientists, the education planners may take advantage of these ‘core strengths’ and initiate programs at the Masters/PhD course level in this direction. As one editorial put it – the time is ripe.

SRINIVASAN RAMACHANDRAN

*National Institute of Allergy and Infectious Diseases,
Building 4, Room 126,
9000 Rockville Pike,
Bethesda, MD 20892, USA*