

Dr. T. MADHAN MOHAN,
Director, Department of Biotechnology,
Govt. of India, New Delhi
E-mail: madhan@dbt.nic.in

BIOINFORMATICS - AN EMERGING TOOL

Its role and relevance to Biotechnology



He is Director for the Biotechnology Information System Network (BTISNET) of the Department of Biotechnology. His doctoral degree was in the areas: Biomedical Engineering (Bio-Electronics), from Indian Institute of Technology, New Delhi and M.Tech from Anna University. Before this he also awarded M.Phil and M.Sc from University of Madras. Since 1989 he is responsible for the implementation of the Bioinformatics culture in the country. He has been developing several databases which are of relevant to Bioinformatics and Biotechnology. He had delivered several popular lectures in Bioinformatics both in India and out side like USA, Taiwan, Malaysia, Israel. He wrote several papers in Bioinformatics and interview for the news papers and TV programs like SUN TV and DD

BIOTECHNOLOGY OVERVIEW

Biotechnology uses living cells and materials to create pharmaceutical, diagnostic, agricultural, environmental, and other products to benefit society. The science of biotechnology is also used to alter genetic information in animals and plants to improve them in some way that benefits people. Because biotechnology essentially uses the basic ingredients of life to make new products, it is both a cutting-edge technology and an applied science. Analysts have predicted that biotechnology will be one of the most important applied sciences of the 21st century.

Ever since man began using microorganisms for the production of fermented food and beverage, Biotechnology has been in existence. About two decades ago, biotechnology became much more of a science (rather than an art). The discovery that strands of DNA called genes give rise to chains of amino acids called proteins, which are important functional components in the cell, had given in our hands the most potential science to benefit mankind. About a decade ago, "protein engineering" became possible as an offshoot of the recombinant DNA technology. Protein engineering differs from "classical" biotechnology in that it is concerned with producing new (man-made) proteins, which have been modified or improved in some way.

In the mid-eighties and early-nineties, it has become possible to transform (genetically modify) plants and animals that are important for food production. "Transgenic" animals, including pigs, cows, sheep, and plants, tomatoes, tobacco, potato, and cotton have now been obtained. Genes introduced may make the organism more resistant to disease, may influence the rate of fruit ripening, or may increase productivity. Recent advances in this area of modern biotechnology are numerous.

NEED OF BIOINFORMATICS FOR THE GROWTH OF BIOTECHNOLOGY

Bioinformatics is an emerging area offering a fundamental tool to the scientific community, particularly to the biologists, to speed up the research, application and commercialization of biotechnology. In fact the best thing, which has happened

Listed below is an overview of recombinant DNA based biotechnology:

- 1953** Double helix structure of DNA is first described by Watson and Crick.
- 1973** Cohen and Boyer develop genetic engineering techniques to "cut and paste" DNA and to amplify the new DNA in bacteria.
- 1977** The first human protein (somatostatin) is produced in a bacterium (*E. coli*).
- 1982** The first recombinant protein (human insulin) appears in the market.
- 1983** Polymerase chain reaction (PCR) technique was conceived.
- 1990** Launch of the Human Genome Project (HGP), an international effort to sequence the human genome.
- 1995** The first genome sequence of an organism (*Haemophilus influenzae*) is determined.
- 2000** The first draft of the human genome sequence is completed.

towards the end of the 20th century and in the 21st century, is the marriage between biotechnologists and information technologists leading to the growth and development of this field. India took a lead in this direction as early as in the 1980s, but more precisely from 1986 onwards by establishing a strong base of Bioinformatics with the necessary infrastructure. As a result one has seen major breakthroughs in biology and the growth of biotechnology has been phenomenal in the last decade, especially with the most outstanding technological breakthrough of the 20th century wherein, a draft of the human genome sequence was completed in 2000 and now the human genetic code is completely available in the database. This has provided the world scientific community information on the vast sequence and structure of the Genomes and the crystal structures leading to increasing dependence on computational approaches. Biotechnology has emerged

as a front-line area with vital significance in unraveling secrets of life, particularly in the studies of new biology and biotechnology. The genomic revolution has underscored the central role of Bioinformatics in understanding the very basic of life processes.

Let us start with the issue of the genetic resources and precious biodiversity of India as also of the world. The need to conserve, protect and utilize them on a sustainable basis is the need of the hour. Genetic resources are used to improve the agricultural productivity and are the vital components of the global environment. Gene Bank concept has become very important, both for ex-situ conservation and the utilization of the valuable germplasm on a long-term basis.

In fact, world over the scientists are associating the computers and the "new biology", looking at a speed with which data collection is done, automated high throughput data gathering, i.e. sequencing the whole genome and the web based data "Warehouses" integrate the information from the research efforts made all over the world. We know that there are more than 3-27 million undiscovered species of plants and animals on the planet earth. However, through the endeavor of the scientists, hundreds of new species are discovered and classified every day. The accelerated progress of information technology has completely transformed the research in life sciences. The focus on genome technologies has made it still more relevant to establish very close linkages between computers and experimental biologists. There are many developments, for e.g., Enterprise Bioinformatics, high system computing, genomic data and data base integration and DNA amplification and sequencing, genome assembly and gene types, web based Bioinformatics, biochips and gene expression analysis and Proteomics and mass spectrometry.

Recent studies have shown that even the implication of mathematical theory in cell biology have become very important. In fact Leroy Hood from the institute for systems biology, in Seattle says "the future will be the study of genes and proteins of organisms in the context of their informational pathways and networks". Similarly the areas of modular biology, synthetic biology is at the forefront of scientific excitement.

The announcement of the remarkable success of the human genome project supported in 1986 and making the draft sequence available to the people was a landmark in the history of modern biology and science. This has generated tools to produce the whole gene catalogues for many microbes, the plant *Arabidopsis thaliana*, the fruit fly *Drosophila melanogaster*, the roundworm *Caenorhabditis elegans*, and soon the puffer fish *Fugu rubripes*. The complete genome sequences have almost served as recipes for life and the foundation of modern biology in the 21st century aimed at the biological goals as defined by the scientific community world over as "Achieve a fundamental, comprehensive and systematic understanding of life".

It is important now that scientists work towards characterization of the full repertoire of molecular machines of the living systems. Also they must understand how these machines are orchestrated with the life system of single cell and complex multicellular organisms.

CONCEPT OF BIOINFORMATICS

Bioinformatics or Life Science Informatics is a young, fast developing science branch, dealing with storage, analysis integration and simulation of molecular biological data. Bioinformatics is an applied discipline that utilizes computational tools to conceptualize biology. It has emerged as a cutting edge technology and a knowledge revolution. Extraordinary growth of information technology and unprecedented advances in molecular biology and recombinant DNA techniques have ushered in the age of Bioinformatics.

Human kind is on the brink of another revolution. There is no doubt that the mapping of the human genome, completed in June 2000, is one of the greatest scientific advancements in history. This breakthrough in biological research was made possible by advancements in using Bioinformatics and computational Biology. Areas such as proteomics, Genomics, combinatorial chemistry, statistics, nanotechnology, spectroscopy and structural and

computational Biology will have increasing applications of Bioinformatics in days to come.

GOAL OF BIOINFORMATICS

The major goal of bioinformatics is to obtain the complete sequences of as many different genomes as possible. By having that sequence information, companies and research organizations can start to do what is commonly referred to as "sequence-based biology." They take that sequence information, and use it to give scientists more direction as to how they should design experiments, and how these scientists should analyze the experiments. So bioinformatics is causing a fundamental shift in how scientists actually approach molecular biology. The "Holy Grail" for bioinformatics is to map all the genes in the human body and decipher, what the role each gene and its associated base pair plays in the expression of the gene.

ROLE AND RELEVANCE OF BIOINFORMATICS IN BIOTECHNOLOGY

Bioinformatics is becoming increasingly important due to the interest of the pharmaceutical industry in genome sequencing projects. There is a vital need to harness this information for medical diagnostic and therapeutic uses, and there are opportunities for other industrial applications. This field is evolving rapidly, which makes it challenging for biotechnology professionals to keep up with recent advancements. The area has evolved to deal with four distinct problems viz. (i) Handling and management of biological data, including its organization, control, linkages, analysis, and so forth. (ii) Communication among people, projects, and institutions engaged in the biological research and applications. Communication includes e-mail, file transfer, remote login, video conferencing, electronic bulletin boards and establishment of web-based information resources; (iii) organization, access, search and retrieval of biological information, documents, and literature; and (iv) analysis and interpretation of the biological data through the computational approaches including visualization, modeling & simulation, and development of algorithms for highly parallel processing of complex biological structures. The following are the thrust areas of

Biotechnology, which are fully dependent on Bioinformatics:

COMPARATIVE GENOMICS:

There is an assumption that the similarity of two sequences, whether it is DNA, RNA or protein implies functional correlation. Some of the most successful bioinformatics applications deal with this kind of analysis. Several software tools have been developed like BLAST that efficiently performs sequence-alignments against large databases of known sequences. It has become a routine task to compare a new sequence against several databases.

FUNCTIONAL GEOMICS

To investigate genes in their cellular context, expression analysis via Microarray and DNA-chips takes place. The comparison of expression patters of well-defined metabolic states allows identifications of pathological phenotypes on a molecular level. Based on the identifications, a reverse engineering of pathways and localization of pharmaceutical targets is possible.

PROTEOMICS

The publication of entire genome sequences led to a shift of interests from pure DNA sequencing to Protein localization and characterization within their cellular context. This became necessary as one gene can give rise to a number of products via mechanisms such as splicing or post-transnational modifications. The proteome refers to the identification and analysis of all the proteins of a cell. This involves the determination of structures as well as the identifications of protein interactions and function in biological pathways.

STRUCTURAL GENOMICS

Structural Genomics covers the calculation of three-dimensional structures based on the sequence of a macromolecule. The theoretical basis of the relationship between sequences and structure is the most fundamental problem of in silico biology. From a computational point of view the so-called folding problem is the most demanding objection of computational biology. Nevertheless only

the knowledge of the structure of a protein can provide a deeper understanding of its function.

PHARMACO GENOMICS

The development of drugs aims to maximize effect and minimize side effects. It would therefore be very convenient to personalize drugs for each patient. The genetic variations among all human is only 0.1% of the total DNA. The differences are mostly point mutations, having phenotypic impact. These so-called SNPs (Single Nucleotide Polymorphisms) become good candidates for drug development and diagnosis. With the investigation of SNPs using Microarrays the investigation of entire populations becomes feasible. Databases of SNPs will reveal the patterns that cause cancer or Alzheimer.

CELLOMICS OR SYSTEMS BIOLOGY

If sufficient data is available and all relevant components for life are identified more complex interactions can be investigated. For a holistic biological understanding of cell, simulations of cells, entire organisms and populations provide new insights. The simulation of life in silico is a future directive for bioinformatics that has started now.

BIOINFORMATICS NETWORK IN THE COUNTRY

DBT'S INITIATIVE:

India's achievements and strategic advantage in information technology and biotechnology have positioned the country to play a crucial role in the post-genomic era. This role would range from analysis of genes and structures, defining metabolic pathways to identification and validation of drug targets and classification of the natural resources based on genetic profiles. Biotechnological applications in agriculture, medicine and environment protection and biodiversity conservation have not only scientific relevance but also offer opportunities for overall sustainable development, most desirable for 21st Century. The countries rich biodiversity and bioresources can reap the harvest from these scientific breaks through.

With the objective of integrating the

country's vast scientific expertise in the different areas in biotechnology and computational sciences, the Department of Biotechnology envisioned in 1986 the establishment of a distributed information network. The network was to provide a common platform for exchange of information among the scientific community and in the process speed up scientific discoveries. The network, referred to as the Biotechnology Information System Network, (BTISnet) presently consists of sixty-one centres spread all over the country. Over the years, the network has created a resource base including computational infrastructure, databases and software. Human resource development through short and long-term programmes, research and applications in frontier areas of Bioinformatics and biotechnology, are the main objectives of the Bioinformatics programme.

The BTISnet has established a link among scientists and organizations involved in research and development activities in Biotechnology. The network today offers a single window information resource in the country covering inter-disciplinary areas of biotechnology and molecular biology. Six National Facilities on Interactive Graphics are dedicated to the promotion of molecular modeling and other related activities. More than 50 subject specific databases have been developed through the BTISnet. Several major International databases for application to genomics and proteomics have been established in the form of Mirror sites as part of the network. These databases are being linked through high speed and large bandwidth network in the form of VPN called BIOGRID INDIA to promote faster sharing of information as well as to encourage tandem research by various R&D labs and industrial units. These sites are designed to act as knowledge pathways for discoveries in biotechnology. At the micro level, the programme deals with the various issues related to the biological data. It covers the development of data analysis tools, modeling of biological macromolecules and their complexes, metabolic pathways, designing of new molecules such as drugs, peptide vaccines, proteins etc.

HUMAN RESOURCE DEVELOPMENT

ADVANCED DIPLOMA COURSE

The skilled human resource in bioinformatics is of utmost importance for effective implementation of the high throughput programmes in biotechnology. The BTISnet is running five long-term advanced diploma courses



in bioinformatics at the post-M.Sc. level through leading universities of the country: Madurai Kamaraj University, Madurai; University of Pune, Pune; University of Calcutta, Jawaharlal Nehru University and Pondicherry University. The number of seats at each University has been increased based on the demand from the industrial sector. Each course is now producing approximately 15-20 candidates per year. Every year more than 60 trained professionals are being generated through this programme.

POST-GRADUATE DEGREE COURSES

Post-Graduate courses such as M.Sc., M.Tech and PhD in bioinformatics have been introduced through Pune University, IIT-Allahabad and JNU, New Delhi, to promote leadership quality human resource in the area of bioinformatics.

SHORT-TERM TRAINING

The BTISnet is also supporting and

organising several short-term training and workshops in front-line areas of bioinformatics and Biocomputing all over the country. A large number of scientists and research scholars from all over the country are making use of these training programmes of which more than 70 have been conducted this year.

R & D ON BIOINFORMATICS

The Bioinformatics Centres are being extensively used for intensive research by the hosts and neighboring institutions. The thrust areas provided are sequence analysis and molecular modeling. Apart from supporting research, the centres are also actively engaged in research on frontier areas of bioinformatics. Intrinsic research in the bioinformatics centres includes gene analysis, protein structure prediction & engineering, modeling macromolecules assembly, evolutionary biology and mechanisms of disease. Under the programme several R&D projects have been supported and the increase of R&D proposals in Bioinformatics has been witnessed in this year. The Department is considering this area as a thrust area to extend more support towards development of Bioinformatics tools.

CONCLUSION

At the end, this is to re-emphasize the importance of knowledge in the field of biology, biotechnology and Bioinformatics. The country today need a strong knowledge base, a fountain of human resources skilled in the multidisciplinary sciences and revival of the prestige in India's efforts as we did long ago involving ancient culture and traditions. As the new field of biology demands the sustainable use and conservation of the precious biodiversity of this county, equally we need to conserve our rich cultural heritage and value system.

