

BIOINFORMATICS: A REVIEW**G. Garg, D. Singh, K. Dashora, Swarnlata Saraf, S. Saraf*, G. Engla¹ and P. Jain¹**

Institute of Pharmacy, Pt. Ravishankar Shukla University, Raipur - 492 010 (India)

¹Sagar Institute of Research & Technology, Pharmacy, Bhopal (India)*(Received February 23, 2005; Accepted May 12, 2005)***ABSTRACT**

The bioinformatics opened new dimensions for the scientists working in biological science and growth of biological information as a result of extraordinary progress in the experimental sciences of molecular biology, structural biology and biotechnology. Bioinformatics is an interface between biological and computational sciences, it is an application of computational techniques, which helps us understand and organize the information associated with biological macromolecules. India is one of the few countries of the world to implement the idea of distributed resources in biotechnology through the biotechnology information system network that currently comprises many bioinformatics centers. Bioinformatics is the convergence of analytical and computational tools with the discipline of biotechnology research and a major contributor to the drug discovery or gene therapy research programs in biotechnology.

INTRODUCTION

The 21st century will be known for the beginning of newer aspects in biological science. The impact although presently in its early stages, will grow in importance, in part, because information technology (IT) will enable scientists to take full advantages of advances in areas such as genomics and proteomics. At the same time, biological systems may provide useful models for advances in computing. Indeed, the information technology and biotechnology revolutions are not likely to continue as separate phenomena.

The term Bioinformatics is recent invention not appearing until around 1991 and then only in context of the emergence of electronic publishing biological data is produced at phenomenal rate. For example as of August 2000, Genbank repository of nucleic acid contained 8,214,000 entries and the SWISS-PORT database of protein sequences contained 88,1661. On an average these databases are doubling in size every 15 months. As a result of this surge in data, computers have become essential to biological research such an approach is ideal because of the ease with which computers can handle large quantities of data and probe the complex dynamics observed in nature. Bioinformatics is often defined as application of computational techniques to understand and organize the information associated with biological

macromolecules, thus the Bioinformatics is interface between biological and computational sciences. In short Bioinformatics is a management information system for molecular biology and has many practical applications.

Bioinformatics: Thrust areas in biotechnology

The following are the thrust areas of biotechnology, which are fully dependent on Bioinformatics

- 1. Genomics:** Genomic studies in Bioinformatics have concentrated on model organisms, and the analysis of regulatory systems. Structural genomics through Bioinformatics assigns structures to the protein products of genome by demonstrating similarity to proteins of known-structure². 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. Bioinformatics help in to investigate genes in their cellular context; expression analysis via micro array and DNA-chips takes places. The comparison of expression patterns of well-defined metabolic states allows identifications of pathological phenotypes on a molecular level.
- 2. Proteomics:** The publication of entire

genome sequences led to a shift of interests from pure DNA sequencing to protein localization and characterization with in their cellular context. This became necessary as one gene can give rise to a number of products via mechanisms such as splicing or post-translational modifications. The proteome refers to the identification and analysis of all proteins of a cell. This involves the determination of protein interactions and function in biological pathways. Protein sequence databases are categorized as primary, composite or secondary.

3. **Structural genomics:** Structural genomics covers the calculation of three-dimensional structures based on the sequence of a macromolecule. The theoretical basis of relationship between sequences and structure is the most fundamental problem in silico biology. From a computational point of view the so called folding is the most demanding objection of computational biology.
4. **Pharmacogenomics:** The development of drugs aims to maximize effects 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 single nucleotide polymorphisms (SNPs) become good candidates for drug development and diagnosis.
5. **Cellomics or System 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, simulation of cells, entire organism and populations provide new insights.

Bioinformatics: Intellectual property

Intellectual property rights (IPR) are the most essential part of today's business, meant for protecting any intangible property. Biotechnology has also attracted intellectual property protection. Bioinformatics is the convergence of computational and analytical tools with the discipline of biotechnology research. Bioinformatics industry is a major contributor to the drug discovery or gene therapy research programs in biotechnology, major areas of Bioinformatics, which need intellectual property protection, are:

- (a) Analytical and information management tools eg: modeling techniques, databases,

algorithms software etc.

- (b) Genomics and proteomics
- (c) Drug discovery/ design.

Examples of IPR are patent, copyright, trademark, geographical indication and trade secret. A patent is an exclusive monopoly granted by the government to an inventor over his invention for limited period of time.

The traditional biotechnology has attracted and still continues to attract intellectual property protection. An early example was the US patent 141072 granted to Louis Pasteur on yeast used in brewing beer. But the modern technologies have attracted most attention in terms of their impact and relationship with the intellectual property system. The decoding of the double helix structure of DNA by Watson and Crick (1953) was a major scientific breakthrough that helped users in the new technology of modern biotechnology. As the functioning of the DNA became better known, subsequent discoveries made possible a major new field of genetic engineering, involving study and alteration of DNA in a living cell, by identifying a gene with useful properties, cloning it and inserting into a host³⁻⁵.

Bioinformatics: Recent developments

The recent developments in Bioinformatics include micro array which are efficient in working and less time consuming and are described as follows:

DNA micro array:

Thousands of genes and their products (e.g. RNA and proteins) in a given living organism functions in a complicated and coordinate way that creates the mystery of life. However traditional methods in molecular biology generally, work on a "one gene in one experiment" basis, which means that the throughput is very limited and the whole picture of gene function is hard to obtain. A new technology, called DNA micro array, has attracted tremendous interests to monitor the whole genome on a single chip so that to obtain the better picture of the interactions among thousands of genes simultaneously. Terminologies that have been used in the literature to describe this technology include biochip, DNA chip, DNA micro array and gene array. The term genome chip is used indicating that this technology is meant to monitor the whole genome on a single chip. Genome chip would also include the increasingly important and feasible protein chip technology. Base pairing is the underlining principle of DNA micro array. An array is an orderly arrangement of samples.

It provides a medium for matching known and unknown DNA samples based on the base pairing rules and automating the process of identifying the unknowns. An experiment with a single DNA chip can provide researchers information on thousands of genes simultaneously - a dramatic increase in throughput.

There are two major application forms for the DNA micro array technology:

1. Identification of sequence (gene/ gene mutation): When the natural cell division of a microbe takes place there are a little probabilities of occurrence of inherited changes. However, still due to some environmental factors some strains exhibit changed characteristics. This whole process termed as mutation.
2. Determination of expression level of genes: Gene medicines are generally based on gene expression systems that contain a therapeutic gene and delivery system. A gene delivery system controls the distribution and access of gene expression unit to the target tissue. Its recognition by cell-surface receptors and its intracellular trafficking⁶.

Applications of DNA micro array technology:

- a. Toxicological research: Toxicogenomics
- b. Disease diagnosis
- c. Drug discovery: Pharmacogenomics
- d. Gene discovery⁷⁻⁹

Role of Bioinformatics

Bioinformatics is becoming increasingly important due to the interest of the pharmaceutical industry in genome sequencing projects. There is a vital need to this information for medical diagnostic and therapeutic uses, and there are opportunities for other industrial application. 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 for distinct problems like

1. Handling and management of biological data, including its organization, control, linkages, analysis, and so forth.
2. Communication among people, projects and institutions engaged in biological research and applications.
3. Organization, access, search and retrieval of biological information, document and literature.
4. Analysis and interpretation of the biological data through the computational approaches

including visualization, modeling, simulation and development of algorithms for highly parallel processing complex biological structures.

Applications of Bioinformatics:

1. **Finding homologues:** The driving force behind Bioinformatics is the search for similarities between different biomolecules. Apart from systemic organization of data, identification of protein homologues has some direct practical uses. The most obvious is the transferring information between related proteins, i.e. given a poorly characterized protein it is possible to search for homologues that are better understood and with caution apply some of the knowledge of the later to the former studies could be made in low level organisms like yeast and the results applied to homologues in higher level organisms such as humans where experiments are more demanding.
2. **Rational drug designs:** The medical applications of Bioinformatics have been aiding rational drug design. Given the nucleotide sequence, the probable amino acid sequences of the encoded protein can be determined using translation software sequence search techniques can then be used to find homologues in model organisms, and based on sequence similarity, it is possible to model the structures.
3. **Large scale censuse:** Although database can efficiently store all the information related to genomes, structures and expression data sites it is useful to condense all this information into understandable trends and facts that users can readily understand. Broad generalization help to identify interesting subject areas for further detailed analysis, and place new observations in a proper context. This enables one to see whether they are unusual in any way¹⁰.
4. **Applications in medical sciences:** Further advancement in Bioinformatics combined with experimental genomics for individual are predicted to revolutionize the further of healthcare. A typical scenario for a patient may start with postnatal genotyping to assess susceptibility or immunity from a

specific disease and pathogen. With this information, a unique combination of vaccine could be prescribed, minimizing the healthcare costs of unnecessary treatments and anticipating the onslaught of diseases later in life.

Future goals

The possible use of bioinformatics in replacement therapy, a defective gene is inserted somewhere in the genome so that its product could replace that of a defective gene. This approach may be suitable for recessive disorders which are marked by deficiency of an enzyme or other proteins.

Bioinformatics can improve the human life improvement of a specific disease *i.e.* introducing a growth hormone gene to increase the height. Bioinformatics help in to investigate gene in their cellular context; this approach can help in injection or insertion of gene into germ cells, *i.e.* into fertilized eggs is known as germ line therapy. Here, the inserted genes would be passed on to future generation too.

Bioinformatics can help in eugenic approach for betterment of human life is brought

about by inserting genes to alter or improve complex traits of a person *i.e.* intelligence. However, it is far beyond the current technological feasibility.

CONCLUSION

With the current deluge of data, computational methods have become indispensable to biological investigations. Originally developed for the analysis of biological sequences Bioinformatics now encompasses a wide range of subject areas including structural biology, genomics and gene expression studies. In particular, we discussed types of biological information and data bases that are commonly used, examined some of the studies that are being conducted with reference to transcription regulatory and finally looked at several practical applications of the field.

Bioinformatics has not only provided greater depth to biological investigations, but added the dimension of breath as well in this way we are able to examine individual systems in detail and also compare them with those that are related in order to uncover common principles that across many systems and highlight unusual features that are unique to some.

REFERENCES

1. Reichhardt, T., It's sink or swim as a tidal wave of data approaches, *Nature*, **399**: 517-20 (1999)
2. Chotia, C. and Gerstein, M., Advances in Structural Genomics, *Curr. Opin. Structu. Boil.*, **27(23)**, 4658-70, (1999)
3. Hulquist, S., Harrison, R. and Yang, Y., Patenting Bioinformatics Inventions; Emerging Trends In The United States, *Nature*, **20**: 320-25, (2002)
4. EFB (European federation of biotechnology) task group on public perception of biotechnology, patenting in biotechnology, briefing paper 1, second edition, (1996)
5. Chow, M. and Fernandez, D., Intellectual property strategy in Bioinformatics, proceedings virtual conference in genomics and Bioinformatics, (2001)
6. Tomlinson, E. and Rolland, A.P., *J. Control Rel.*, **39**: 357-72, (1996)
7. Ekins, R., Micro Arrays; Their Origins And Applications, *Trends in Biotech.*, **17**: 217-18 (1999)
8. Giot, L., Knight, J.L. and Judson, R.S., A Comprehensive Analysis of Protein-Protein Interactions, *Nature*, **403**: 623-627, (2000)
9. Perou, C.M., Ross, D.T., Shalon, D. and Brown, P.O., Distinctive Gene Expression Patterns In Human, *Proc. Natl. Acad. Sci.*, **16**: 9212-17, (1999)
10. Luscombe, N.M., Greenbaum, D. and Gerstein, M. What is Bioinformatics- An Introduction, Yearbook of Medical Informatics, Oxford University Press, Oxford, 83-96 (2001)